

# Solving a 3D structural puzzle

**Hoeck, Casper**

*Publication date:*
2016

*Document Version*
Publisher's PDF, also known as Version of record

# Solving a 3D structural puzzle

PhD Thesis

Casper Hoeck


Department of Chemistry

Technical University of Denmark

May 2016

# Solving a 3D structural puzzle

Casper Hoeck

# Preface

The work presented in this thesis was carried out during my three years of research as a PhD student at the Department of Chemistry, the Technical University of Denmark. The work was supervised by Associate Professor Charlotte Held Gotfredsen, and funded by the Department of Chemistry. A three-month external stay was conducted in the group of Dr. Craig Butts at the School of Chemistry, University of Bristol, UK.

The thesis is divided into nine chapters, covering three main topics. The first topic is nuclear Overhauser effect correlations and $^{3}J_{HH}$-coupling constants in 3D structural investigations of organic compounds (Chapter 3). This is followed by a chapter concerning structural information from long-range coupling constants and new pulse sequences to determine these (Chapter 4). The final topic covers multiple chapters and focuses on 3D structural information from residual dipolar coupling constants, from increasing the number of available inter-nuclear vectors by long-range coupling constants (Chapter 6) to enantiodiscrimination (Chapter 7), and a chapter on structural flexibility and a new method of back-calculating (Chapter 8).

First and foremost, I would like to thank Lotte for accepting me as a PhD student in her group, for giving me an opportunity to spend three years on exciting research within NMR spectroscopy and for her guidance and support throughout. It has been a privilege, and I have learned a lot in the three years. I would like to thank the entire NMR group, especially to former group member Louise Kjærulff for great NMR talks, discussions and cooperation on new NMR experiments, and Anne Hector for always being very helpful. Regarding NMR experiments I would also like to thank Ole W. Sørensen for invaluable help on theoretical NMR spectroscopy in our collaborations. Many thanks to Sebastian Meier for always being helpful and especially for proofreading my thesis. For this I also owe Niclas Hoeck and Pernille Rønn-Nielsen my thanks. Thanks to the people of building 201 (now 211) and the rest of the Department of Chemistry, for three highly enjoyable and memorable years.

I am thankful that I through my studies have had extraordinary collaborators: Christian A. Olsen, KU, and former group members Alex Maolanon and Jesper Villadsen with whom I was involved in work regarding the 3D structures of azumamides, and Morten Meldal, KU and Ming Li with whom we collaborated on projects regarding NMR investigations of molecular recognition of synthetic peptides. I would also like to thank Thomas Ostenfeld Larsen, from Systems Biology (DTU) and Thomas Isbrandt for our continued collaboration on secondary metabolites from *A. homomorphus* as well former group member Lene Pedersen for shared work on 3D structures of metabolites from other fungi.

I am grateful to Craig hosting me as a visiting scholar, and the entire NMR group at the School of Chemistry, University of Bristol, (Craig, Jess, Iyke, Claire, Tom, Siying and Paul) for being very welcoming and making my three-month external stay enjoyable and rewarding.

Last, but not least, my deepest thanks to my family, my friends, and Pernille, for unwavering encouragement and support.

Casper Hoeck, May 31st 2016, Kgs. Lyngby

# Abstract

Nuclear magnetic resonance (NMR) spectroscopy is a versatile tool in analytical chemistry, highly suitable for structural elucidation of organic molecules – as well as multiple other areas of research. The subjects covered within this thesis all concern methods which allow a shift from covalent to spatial structural information using NMR spectroscopy.

Experimental distances from nuclear Overhauser effect (NOE) correlations, and dihedral angles from $^{3}J_{HH}$-coupling constants, were used to obtain 3D structural information for several natural and synthetic compounds. The stereochemistry of novel natural compounds was determined, including that of a bicyclic non-ribosomal peptide (a novel structural motif), a steroid and several polyketides. Structural insights were gained for potential anti-cancer agents; the azumamides, including synthetic analogues. Differences in the conformational space of solution state compounds were identified experimentally between structural analogues, and compared to the *in vitro* potency of the compounds. The structures of two peptides that exhibited a high degree of molecular recognition were investigated, resulting in the elucidation of a possible mode of interaction. Also a major assumption in the calculation of distances from NOEs, the assumption of equal rotational correlation times between proton pairs, was investigated for molecules in organic solvents.

Two spin-state selective ($S^3$) HMBC experiments were developed for measurements of homonuclear and hetereonuclear long-range coupling constants, respectively. The new NMR experiments were based on two existing experiments, the multiplicity edited HMBC and the HAT HMBC, which were combined to obtain $S^3$ editing of long-range homonuclear coupling constants. The output of the first $S^3$ HMBC experiment was HMBC type spectra with $^{n}J_{CH}$ correlated cross-peaks, from which $^{n+1}J_{HH}$-coupling constants were sign-selectively determined with high accuracy. Very small coupling constants, including previously unreported coupling constants from strychnine, were extracted, with all experimental values correlating very well to theoretical coupling constants from DFT calculation. A pulse segment was developed to change the polarization of the CH-H pairs in the homonuclear $S^3$ HMBC, to gain $S^3$ edited $^{n}J_{CH}$-coupling constants in the cross-peaks. While only determining coupling constants to methine carbons, the extracted experimental coupling constants correlated very well to theoretical coupling constants, thus extending the $S^3$ HMBC methodology to include both $^{n+1}J_{HH}$- and $^{n}J_{CH}$-coupling constants.

Residual dipolar couplings (RDCs) are a relatively late addition to the small molecular NMR community, where alignment media are used to obtain anisotropic samples, which allows for RDCs to be extracted. The number of internuclear vectors for the correlation of RDCs to 3D structures is often limited for small molecules. Homonuclear RDCs were extracted by use of the homonuclear $S^3$ HMBC that correlated well to alignment tensors from $^{1}D_{CH}$-coupling constants, thus increasing the number of inter-nuclear vectors. The topic of enantiodiscrimination by RDC measurements of rigid organic molecules was also investigated, and new alignment media were developed to allow slight discrimination of enantiomers by stretched polymers. Finally a new method of back-calculation of RDCs from 3D structures was developed and tested, which copes better with multiple conformers than the commonly used SVD methodology. The approach thus resulted in good conformer populations for several small molecules, including multiple cinchona alkaloids.

# Resumé

NMR (nuclear magnetic resonance) spektroskopi er et alsidigt og vigtigt værktøj i analytisk kemi, der ofte er involveret i strukturopklaring af organiske molekyler såvel som i mange andre forskningsområder. Denne afhandling omhandler metoder til at opnå rummelig information om molekylære strukturer ved hjælp af NMR spektroskopi.

3D strukturer af nye natur- og syntetiske stoffer blev således undersøgt ved hjælp af NOE (nuclear Overhauser effekt) korrelationer og tre-bindings *J*-kobling konstanter. NOE korrelationer kan bruges til at bestemme afstande mellem atomer i de organiske strukturer, og *J*-koblingskonstanter kan give information of den dihedrale vinkel i H-X-X-H systemer (hvor X oftest er C, N, O). For naturstofferne, inklusivt et stof med et hidtil ukendt to-rings-system, ledte NMR data til, at alle relative stereocentre blev bestemt. NOE og *J*-koblinger resulterede også i ny viden om en gruppe af potentielle anti-cancer stoffer, kaldet azumamiderne, hvor et naturstof og syntetiske analoger blev undersøgt. Der blev identificeret forskelle i det konformatielle rum mellem de forskellige strukturer, hvilket blev relateret til deres anti-cancer aktivitet. I et andet projekt blev interaktionen mellem to peptider undersøgt, via viden om deres respektive struktur i vand fra afstande mellem atomer. De to peptiderne er af særlig interesse, da de selektivt interagerer med hinanden i vandig opløsning, ligesom f.eks. proteiner i celler. Derudover blev det undersøgt, i hvor høj grad proton-par i organiske strukturer roterer og bevæger sig identisk, hvilket er en antagelse, der gøres for at kunne finde afstande fra NOE korrelationer. Det undersøgtes også om det var muligt a korrigere afstandene, fra viden om proton-pars relative rotationshastigheder i organiske solventer.

To nye NMR eksperimenter, kaldet $S^3$ (spin-state selective) HMBC, blev udviklet til at måle henholdsvis proton-proton og proton-kulstof *J*-koblingskonstanter i organiske strukturer, hvor de to koblende atomer er adskilt af op til flere kemiske bindinger ($^{n}J_{CH}$ og $^{n+1}J_{HH}$, n=2-4). De to NMR eksperimenter er baseret på to eksisterende eksperimenter, der kombineredes sådan, at det var muligt at måle proton-proton koblingskonstanter mellem en proton bundet til et kulstof og en proton, der kobler til dette kulstof. En ny pulssekvens blev tilføjet dette første $S^3$ HMBC eksperiment, hvilket gjorde det muligt også at måle proton-kulstof koblingskonstanterne. De målte koblingskonstanter fra begge sekvenser korrelerede i meget høj grad med teoretisk bestemte koblingskonstanter, selv når disse var meget små. Dette medfører derfor muligheden for øget strukturel viden.

Residuale dipolære koblinger (RDCs), en forholdsvis nyligt introduceret parameter for små molekyler, og er baseret på vinklen mellem en internukleær vektor og det magnetiske felt fra NMR spektrometret. Her er få internukleære vektorer ofte et problem for små molekyler. Det blev vist at det homo-nukleare $S^3$ HMBC kunne benyttes til at bestemme adskillige RDCs mellem protoner, og derved blev flere internukleære vektorer bestemt. For at kunne aflæse RDCs bruges såkaldte alignment media, hvor et overskud af en bestemt orientering af molekylerne i NMR prøven skabes. Nye alignment media blev syntetiseret, og disse muliggjorde bestemmelsen af enantiomerer (spejlede molekyler) ud fra deres aflæste RDCs. Endeligt blev en ny beregningsmetode af RDCs fra 3D strukturer testet. Metoden fungerede klart bedre end standardmetoden SVD, når fleksible strukturer blev undersøgt, hvilket blev vist for flere molekyler.

# Publications

**I** Maolanon, Alex R; Villadsen, Jesper S; Christensen, Niels J; **Hoeck, Casper**; Friis, Tina; Harris, Pernille; Gotfredsen, Charlotte H; Fristrup, Peter; Olsen, Christian A. Methyl Effect in Azumamides Provides Insight Into Histone Deacetylase Inhibition by Macrocycles. *Journal of Medicinal Chemistry*, **2014**, *57*, 9644-9657.

**II** Petersen, Lene M; **Hoeck, Casper**; Frisvad, Jens C; Gotfredsen, Charlotte H; Larsen, Thomas O. Dereplication Guided Discovery of Secondary Metabolites of Mixed Biosynthetic Origin from *Aspergillus aculeatus*. *Molecules*, **2014**, *19*, 10898-10921.

**III** Kjaerulff, Louise; Benie, Andrew J; **Hoeck, Casper**; Gotfredsen, Charlotte H; Sørensen, Ole W. $S^3$ HMBC: Spin-State-Selective HMBC for accurate measurement of homonuclear coupling constants. Application to strychnine yielding thirteen hitherto unreported $J_{HH}$. *Journal of Magnetic Resonance*, **2016**, *263*, 101-107

**IV** Li, Ming; **Hoeck, Casper**; Schoffelen, Sanne; Gotfredsen, Charlotte H; Meldal, Morten. Specific Electrostatic Molecular Recognition in Water. *Chemistry - A European Journal*, **2016**, *22*, 7206-7214.

**(V)** **Hoeck, Casper**; Petersen, Lene M; Frisvad, Jens C; Larsen, Thomas O; Gotfredsen, Charlotte H. Homomorphosins A-F, novel diketopiperazines. (Draft)

**(VI)** **Hoeck, Casper**; Gotfredsen, Charlotte H; Sørensen, Ole W. $S^3$ HMBC *hetero*: Spin-State-Selective HMBC for accurate measurement of heteronuclear coupling constants. (Draft)

# Abbreviations

| | |
|---|---|
| 1D | One dimensional |
| 2D | Two dimensional |
| 2F | Two-field |
| 3D | Three dimensional |
| AIBN | Azobisisobutyronitrile |
| AP | Anti-phase |
| APhES | 2-acrylamido- 2-phenylethane-1-sulfonic acid |
| AU | Automation |
| ax | Axial |
| Azu A | Azumamide A |
| B3LYP | Becke Three Parameter Hybrid Functionals |
| C10-Phe | Decyl-acryloyl-L-phenylalaninate |
| $CHCl_3$ | Chloroform |
| CM-A | Cyclomorphosin A |
| CM-B | Cyclomorphosin B |
| CORMA | Complete Relaxation Matrix Analysis |
| COSY | Correlation spectroscopy |
| CSGT | Continuous set of gauge transformation |
| CW | Continuous wave |
| DCM | Dichloromethane |
| Des C | Desmethyl-azumamide C |
| DFT | Density Functional Theory |
| DMSO | Dimethyl sulphoxide |
| DMAA | (N,N)-dimethyl acrylamide |
| DQ | Double quantum |
| DQF | Double quantum filtered |
| DSO | Diamagnetic spin orbit |
| DTU | Technical University of Denmark |
| EGDMA | Ethylene glycol dimethylacrylate |
| Epi E | Epi-azumamide E |
| eq | Equatorial |
| ESI | Electrospray ionization |
| EtOAc | Ethyl acetate |
| F1 | Indirect dimension |
| F2 | Direct dimension |
| FC | Fermi contact |
| FDA | Food and drug administration |
| FDAA | 1-Fluoro-2,4-dinitrophenyl-5-L-alanineamide |
| FF | Force field |
| FID | Free induction decay |
| FT | Fourier transform |
| GCB | generalized cosine β |
| GIAO | Gauge invariant atomic orbitals |
| GTO | Gaussian type orbital |
| HAT | Homonuclear J-attenuated |
| HDAC | Histone deacetylase |
| HLA | Haasnoot, De Leeuw and Altona (or Altona) |
| HMBC | Heteronuclear multiple bond correlation |
| HSQBC | Heteronuclear single quantum multiple bond connectivity |
| HSQC | Heteronuclear single quantum coherence |
| INEPT | Insensitive nuclei enhanced by polarization transfer |
| IP | In-phase |
| IPAP | In-phase/anti-phase |
| IPC | Isopinocampheol |
| IRA | Initial rate approximation |
| ISPA | Isolated spin-pair approximation |
| KU | University of Copenhagen |
| L7 | Ligand 7 |
| LC | Liquid crystal |
| LC | Liquid chromatography |
| MA | Mass spectrometry |
| MAE | Mean absolute error |
| MARDIGRAS | Matrix Analysis of Relaxation for discerning the Geometry of an Aqueous Structure |
| MCMT | Multi conformer multi tensor |
| MCST | Multi conformer single tensor |
| MD | Molecular dynamics |
| MeOH | Methanol |
| MM | Molecular mechanics |
| NMR | Nuclear magnetic resonance |
| NOE | Nuclear Overhauser effect |
| NOE/ROE | NOESY/ROESY comparison |
| NOESY | Nuclear Overhauser effect spectroscopy |
| NRP | Non-ribosomal peptide |
| PANIC | Peak Amplitude Normalization for Improved Cross-relaxation |
| PBLG | Poly-γ-benzyl-L-glutamate |
| PDMAA | Poly-(N,N)-dimethyl acrylamide (crosslinked) |
| PEGA | Poly(acryloyl-bis-(aminopropyl)polyethylene glycol) |
| PhEtN | 1-Phenyl ethylamine |
| PhEtO | 1-Phenyl ethanol |
| ppm | Parts per million |
| PSO | Paramagnetic spin orbit |
| *Q* | *Q-factor* |
| RDC | Residual dipolar coupling constant |
| RF | Radio frequency |
| RMSD | Real mean square deviation |
| ROESY | Rotating frame nuclear Overhauser effect spectroscopy |
| *S*/*N* | Signal-to-noise |
| $S^3$ | Spin-state selective |
| SAG | Strain-induced alignment in a gel |
| SD | Spin-dipole operator |
| SM | Secondary metabolite |
| SO | Spin-orbit coupling |
| SP | Stretched polymer |
| SQ | Single quantum |
| STO | Slater-type orbital |
| SVD | Singular value decomposition |
| $T_1$ | Longitudinal relaxation |
| $T_2$ | Transverse relaxation |
| T20 | Target 20 |
| TEA | Triethyl amine |
| TEG-Phe | (Triethylene glycol methyl ether)-acryloyl-L-phenylalaninate |
| TMS | Tetramethylsilane |
| TOCSY | Total correlation spectroscopy |
| UHPLC | Ultra high pressure liquid spectroscopy |
| UV | Ultraviolet |
| YES | Yeast extract sucrose agar |
| ZQ | Zero quantum |

| | |
|---|---|
| $B_0$ | Magnetic field |
| $\omega$ | Frequency in rad |
| $\gamma$ | Gyromagnetic ratio |
| $\hbar$ | Reduced Planck constant |
| $\mu_0$ | Vacuum permeability constant |
| $\sigma$ | Cross-relaxation rate constant |
| $\rho$ | Dipolar longitudinal relaxation rate constant |
| $\eta_{IS}$ | Cross-peak intensity |
| $\eta_I$ or $\eta_S$ | Diagonal/irradiated peak intensity |
| $\tau_c$ | Rotational correlation time |
| $\tau_s$ | Relative intramolecular movements |
| $\tau_o$ | Overall rotational correlation time |
| $\tau_m$ | Mixing time |
| $J$ | $J$-coupling constant |
| $D$ | Dipolar coupling constant |
| $T$ | Total coupling constant |
| $A_a$,$D_a$ | Axial component, degree of alignment |
| $A_r$,$D_r$ | Rhombic component |
| $R$ | Rhombicity |
| $D_{max}$ | Maximum possible dipolar coupling constant |
| $D_{exp,max}$ | Maximum experimental RDC |
| $D_{exp,min}$ | Minimum experimental RDC |

# Content

# 1 NMR spectroscopy: Past to present

The foundation for modern day NMR spectroscopy was laid more than 70 years ago, in 1946 by the work of the groups of Bloch and Purcell who first reported on NMR in bulk materials. Work that awarded them a shared Nobel Prize in Physics in 1952.[1,2] One could look back even further to the introduction of the concept of nuclear spin, the phenomenon exploited in NMR spectroscopy, which was described in the 1920s.[3,4] Since the early discoveries much has happened, from the introduction of superconducting magnets and development of high field instruments, to proceedings in methods of data acquisition, leading to today where NMR spectroscopy is one of the most widely used spectroscopic techniques in chemistry, structural biology and material sciences - and arguably the most information rich when used appropriately. It is based on an electron density dependent energy difference of nuclear spin orientation in an applied magnetic field, as outlined in equation (1.1), and the ability of interacting with these spins using radio frequency (RF) energy.

$$\omega = \gamma B_0(1 - \sigma) \quad \textbf{(1.1)}$$

Where $\omega$ is the resonance frequency of a given nucleus, $\gamma$ is the gyromagnetic ratio of said nucleus, $B_0$ is the applied magnetic field and $\sigma$ is the shielding constant which is dependent on the electron density and configuration of electrons around the nucleus.[3]

In the earliest days of NMR spectroscopy the technique used was continuous-wave spectroscopy, where the frequency of the RF pulse was varied, and the spectra recorded by scanning the spectral range from start to finish.[3,4] This changed in the mid-1960s with the introduction of Fourier transform (FT) techniques by Ernst and Anderson; work for which Ernst was rewarded with the 1992 Nobel Prize in Chemistry.[3,5] FT is a method to extract frequencies from free induction decay (FID), and thus data for the full spectral range may be acquired from a single short RF pulse. NMR spectroscopy has since evolved into a versatile field, largely in part of ever evolving methods of using RF pulses to influence targeted nuclei and output a diverse range of spectra, each with a specific purpose. The large array of pulse sequences, which each hold the promise of a vast amount of information, means that NMR is applicable when investigating very different systems; small molecules, macromolecules and materials in liquid- or solid state alike.

On the hardware side, the introduction of superconducting magnets in the mid-1960s laid the ground for modern day spectrometers.[3,4] Until then 100 MHz was the maximum field strength obtainable, but the development of higher field instruments were available through the new technology, from 220 MHz in 1966, 500 MHz in 1978, 800 MHz in 1995 to modern day instruments pushing the 1 GHz boundaries and beyond. Due to the correlation of magnetic field strength, resolution and signal-to-noise (*S/N*), the shift to bigger magnets has vastly increased the scope of the technology, hand in hand with the development of better probes and other relevant hardware.

This thesis will primarily focus on unique possibilities of NMR spectroscopy that sets the method apart from many other spectroscopic methods, namely observable interactions between nuclei. These magnetic dipole interactions are used extensively in this thesis and are the focus of the theory section.

# 2 Theory – 3D structural information from NMR, part 1

The various projects included in this thesis share a common trait – they all rely on information obtained from magnetic dipolar interactions between nuclei in organic molecules.

The main focus will be on NMR observables which are useful in elucidations and calculations of 3D structures, namely nuclear Overhauser effect correlations (NOEs) (dipolar relaxation, through space), *J*-coupling constants (indirect dipole-dipole couplings, through bonds) and residual dipolar coupling constants (RDCs) (direct dipole-dipole couplings, through space). The three subjects are conceptually depicted in Figure 2.1 and each topic will be treated independently in this chapter and Chapter 5.

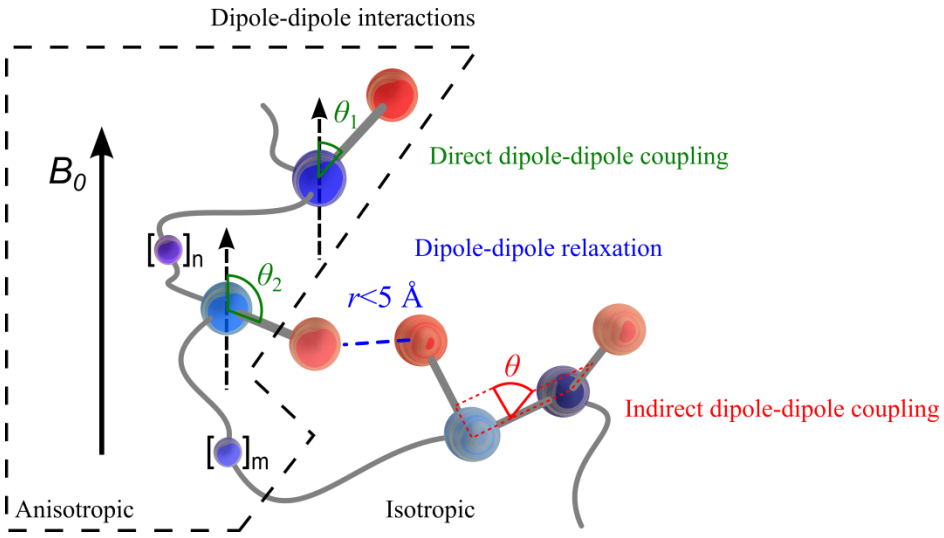



**Figure 2.1. Concepts of magnetic dipole-dipole interactions and the obtainable 3D structural information. In green are the direct dipole-dipole couplings, which may be used to gain information of angles and distances of internuclear vectors independent on their situation in the molecular structure. In blue is the dipole-dipole relaxation, which may give access to local internuclear distances through space. Indirect dipole-dipole couplings are highlighted in red which, among multiple uses, hold local information of the relative position of nuclei, mediated through the bonds in the structure. It is indicated whether the information may be extracted from spectra of molecules under isotropic conditions or if anisotropic conditions are also needed.**

Chosen theory for the different topics is presented, and focus will be on the theory which was applied, either directly or implicitly. This selection results in some rather broad jumps in the presented theory and further information may be found in more comprehensive literature sources on the subjects, e.g. from the cited references.

## 2.1 The nuclear Overhauser effect

The mechanism involved in nuclear Overhauser effect (NOE) spectroscopy is dipolar cross-relaxation of nuclei in close spatial proximity. Being facilitated entirely through space in contrast to e.g. the bond mediated *J*-coupling constants, *vide infra*, NOEs leads to purely spatial information of the nuclei, as illustrated in Figure 2.2.[6,7]

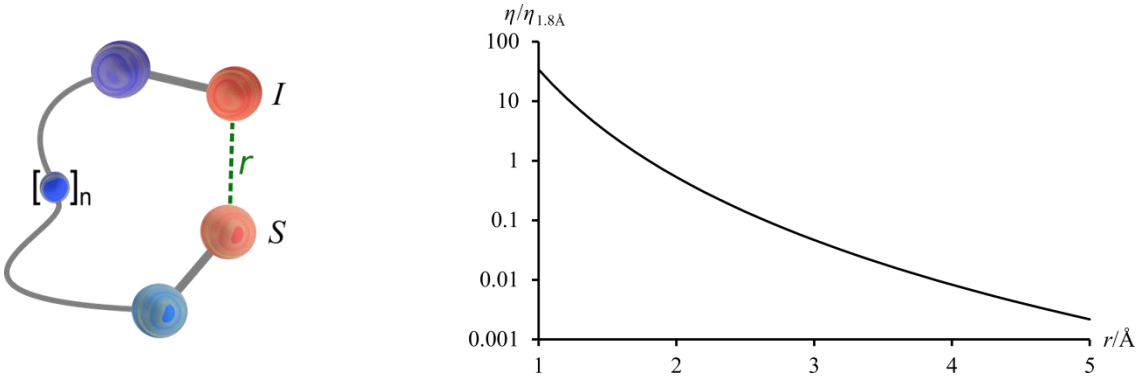



**Figure 2.2. Left: Illustration of the through-space dependance of NOE relaxation of protons (red), blue may be carbon or heteroatoms. Right: Correlation of theoretical relative NOE intensity and distance (in Å) between protons. Relative to a reference distance of 1.8 Å set to 1. Note the logarithmic scale of the y-axis.**

The phenomenon is named after Overhauser, who first proposed that polarization of metal nuclear spins would be observable upon saturation of the metal electrons in 1953.[8,9] Solomon established the theoretical basis and first experimental evidence of NOEs two years later. The proposed Solomon equation, equation (2.1), is thus considered the basis of NOE theory.[7,10] While the NOE effect is equally important in homo- and heteronuclear NMR, the latter exemplified from e.g. NOE enhancement of $^{13}C$ resonances, the homonuclear variant between protons is a prime source of structural information and is the focus of this section. Note that the involved nuclei will be labeled as I and S, even though both are protons (in some publications $I_1$ and $I_2$ or other notation is used for homonuclear cases).

### 2.1.1 Relaxation

The generation of NOE correlations is a relaxation process and some points regarding general relaxation processes in liquid NMR spectroscopy will be made, while the literature provide a much more in depth description.[7] Relaxation in NMR is the mechanism of returning to equilibrium after the spin population has been perturbed by RF pulses and is divided into longitudinal ($T_1$) and transverse ($T_2$) relaxation. In short $T_1$ relaxation is the recovery of the original longitudinal magnetization, and $T_2$ is the loss of transverse magnetization.[7] For small organic molecules, which are the focus of this thesis, the timeframe of $T_2$ and $T_1$ relaxation processes are usually of equal orders of magnitude, while $T_2$ may be much shorter than $T_1$ for macromolecules or solids.[11]

The longitudinal relaxation is the main interest for understanding NOE relaxation. The relaxation process is not caused by spontaneous emission, as this is slow for the energy-levels involved in NMR spectroscopy. Instead it is facilitated by the translations, rotation and internal motions of the molecules in the sample – termed “the lattice”.[7] This may describe the difference between the relaxation properties of different sizes of molecules, as the motions, or tumbling, will be quite different. The energy from the relaxation is taken up by the lattice as the near infinite degrees of freedom of the full lattice translate to the energy levels of the lattice being a continuum. This means that any NMR transition will have a matched energy level in the lattice.

The relaxation mechanism of the NOE relaxation is dominated by the contribution from the relative spin dipole reorientation of neighboring nuclei under rapid isotropic tumbling, as the population of one spin changes when the population of another nuclear spin is perturbed.[6,7] This is the cause of

the dependence of NOEs on the rotational correlation time ($\tau_c$). The initial population shift of a nuclear spin can occur by continuous saturation of a resonance, as illustrated in Figure 2.3, the basis of steady-state NOEs. The continuous saturation effectively leads to equal nuclear spin population differences across the corresponding $\alpha/\beta$ transitions. The NOE is given by the $W_2$ and $W_0$ cross-relaxation where both spins switch spin state simultaneously, see the legend of Figure 2.3 for more information.

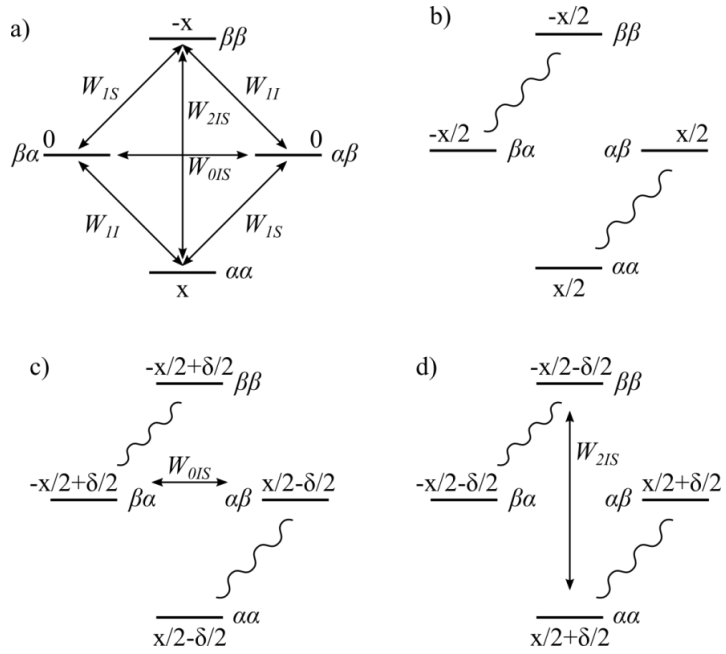



**Figure 2.3. Energy diagrams for the NOE in a homonuclear two-spin system of the protons *I* and *S*. The intensity of proton *I* is proportional to ($N_{\alpha\alpha}$ - $N_{\beta\alpha}$) + ($N_{\alpha\beta}$ - $N_{\beta\beta}$). Inspired by Neuhaus and Williamson.[7]**
**a) The possible spins states, given as I(α)S(β), and transition probabilities (*W*). The $W_2$ transition (the double quantum) and the $W_0$ (zero quantum) transition are the relevant cross-relaxation pathways. The initial populations are $N_{\alpha\alpha}$ = x, $N_{\beta\beta}$ = -x and $N_{\alpha\beta}/N_{\beta\alpha}$ = 0.**
**b) Saturation of the S resonance leads to $N_{\alpha\alpha} = N_{\alpha\beta}$ and $N_{\beta\beta} = N_{\beta\alpha}$. The saturation is maintained throughout.**
**c) The effect of $W_0$ cross-relaxation leads to a transfer of the population δ/2 from $N_{\alpha\alpha}$ to $N_{\beta\beta}$. This leads to a shift towards equilibrium and thus a decrease in the intensity of resonance *I*; a negative NOE enhancement.**
**d) The effect of $W_2$ cross-relaxation leads to a transfer of the population δ/2 from $N_{\beta\beta}$ to $N_{\alpha\alpha}$. This leads to a shift away from equilibrium and thus an increase in the intensity of resonance *I*; a positive NOE enhancement.**

The steady state situation in Figure 2.3 is described by the Solomon equation (2.1), which defines the change of $I_z$ (a vector equal to the population difference between the spin states) over time as a function of the relaxation pathways and the initial population differences $I_z^0/S_z^0$. $\partial I_z$/dt and $S_z$ are equal to 0 for steady state.[7]

$$\frac{\partial I_z}{\partial t} = -(I_z - I_z^0)(W_{0IS} + 2W_{1I} + W_{2IS}) - (S_z - S_z^0)(W_{2IS} - W_{0IS}) \quad \textbf{(2.1)}$$

This equation may be rewritten to describe the NOE enhancement, $f_I\{S\}$, which describes the change of intensity for spin *I* when saturating spin *S*, using that $S_z^0 = (\gamma_S/\gamma_I)I_z^0$.[7]

$$f_I\{S\} = \frac{(I_z - I_z^0)}{I_z^0} = \frac{\gamma_S}{\gamma_I}\frac{(W_{2IS} - W_{0IS})}{(W_{0IS} + 2W_{1I} + W_{2IS})} = \frac{\sigma_{IS}}{\rho_{IS}} \quad \textbf{(2.2)}$$

Where $\sigma$ is the cross-relaxation and $\rho$ is the dipolar longitudinal relaxation rate constant. The cross-relaxation rate determines the intensity of the cross-peaks and is of major interest. It is positive if

$W_{2IS}$ dominates $W_{0IS}$, and negative in the opposite scenario, which is in correlation to what may be found from Figure 2.3. The transition probabilities, or spectral densities, may be calculated from the dipolar interaction Hamiltonian to give the formula for cross-relaxation shown in equation (2.3),[7] where it is utilized that the transition probabilities are proportional to the chemical shift frequencies of the nuclei as the $\alpha\beta\leftrightarrow\beta\alpha$ transition corresponds to a frequency of $|(\omega_I-\omega_S)|$ and the $\alpha\alpha\leftrightarrow\beta\beta$ transition corresponds to a frequency of $(\omega_I+\omega_S)$, which may also be realized from Figure 2.3.

$$\begin{aligned}\sigma_{IS} &= W_{2IS} - W_{0IS} = \frac{1}{10}K^2\tau_c\left(\frac{6}{1+(\omega_I+\omega_S)^2{\tau_c}^2} - \frac{1}{1+(|\omega_I-\omega_S|)^2{\tau_c}^2}\right) \\ &= \frac{1}{10}K^2\tau_c\left(\frac{6}{1+(\omega_I+\omega_S)^2{\tau_c}^2} - 1\right) \\ K &= \left(\frac{\mu_0}{4\pi}\right)\hbar\gamma_I\gamma_S r^{-3}\end{aligned} \quad \textbf{(2.3)}$$

Where $\gamma$ is the gyromagnetic ratio of the nuclei and $r$ is the distance between the involved nuclei, $\mu_0$ is the vacuum permeability constant and $\hbar$ the reduced Planck constant. Equation (2.3) exploits that the difference between $\omega_I$ and $\omega_S$ is much lower than one, especially in the narrow chemical shift range of protons. In the same manner the dipolar longitudinal relaxation rate may be calculated.[7]

All of these equations are based on the steady state experiment which is seldom used today, but more easily explained. The more utilized experiments to obtain NOE spectra is based on inversion, where nuclear spin-population differences are inverted across the transitions, which leads to kinetic NOEs. The method is the most frequently utilized due to fewer artifacts.[7] The kinetic NOE is used to determine the rate of which the steady state is reached instead of performing measurements on the steady state itself, and is implemented as either the truncated NOE (TOE) or as the transient NOE.[7] In transient NOE experiments, which are used throughout this thesis, the resonances are initially inverted and the NOE evolves without RF irradiation for a fixed period, termed the mixing time $\tau_m$. After the evolution an observe pulse is applied to sample the population distribution and obtain the NOE spectra. In this type of NOESY experiment the NOE enhancements initially build up linearly, reach a maximum where linearity is lost and then decay back to zero.

### 2.1.2 The ROESY experiment

The rotating frame NOE (ROESY) experiment, initially dubbed CAMELSPIN (cross-relaxation appropriate for mini-molecules emulated by spin-locking) is a related experiment to the NOESY experiments which evolves the enhancements between elements of transverse magnetization instead of longitudinal magnetization.[7,12–14] For this purpose, a spin-lock from continuous wave (CW) or a pulse train of multiple 180° pulses (T-ROESY) is used for a fixed period of time. The spin-lock period is similar to the mixing time of the transient NOESY experiments. This is advantageous for small molecular NMR, as the ROESY variant of the NOESY longitudinal cross-relaxation depends differently on the rotational correlation time as seen in given in equation (2.4) for CW-ROESY and (2.5) for T-ROESY. [7,12,13,15]

$$\sigma_{IS} = \frac{1}{10} K^2 \tau_c \left( \frac{3}{1 + \omega^2 {\tau_c}^2} + 2 \right) \quad \textbf{(2.4)}$$

$$\sigma_{IS} = \frac{1}{20} K^2 \tau_c \left( \frac{6}{1 + 4\omega^2 {\tau_c}^2} + \frac{3}{1 + \omega^2 {\tau_c}^2} + 1 \right) \quad \textbf{(2.5)}$$

For small- and medium size molecules at a certain rotational correlation time, the NOE enhancement is zero, which is avoided in the ROESY experiments as seen in Figure 2.4 for NOE and T-ROE.

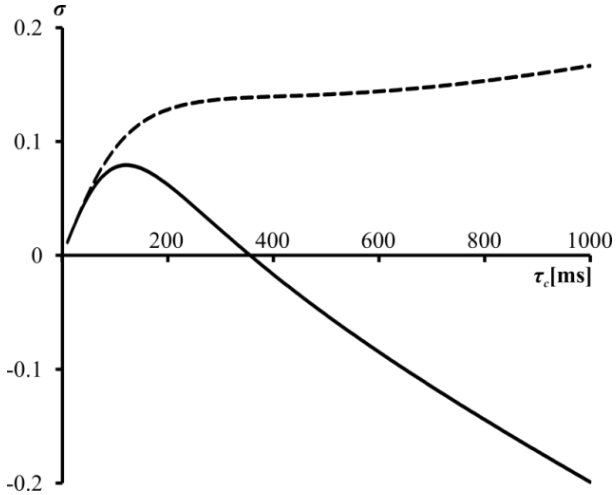



**Figure 2.4. The theoretical cross-relaxation rate ($\sigma$) of NOE (—) and T-ROE (- -) experiments at 500 MHz, for a distance of 2.5 Å. Note that the crossover point is dependent on the field strength. The cross-relaxation rate is shown in contrast to the oft depictured maximum homonuclear enhancement, as the cross-relaxation rate is the actual observable parameter in NOE experiments. Calculated using equations (2.3) and (2.5).**

While there are clear cases where the utilization of ROESY experiments are advantageous due to a better signal to noise ratio, ROESY spectra suffer from TOCSY transfers and a offset-dependency of the spin-lock, which needs to be accounted for.[7,12] This makes the ROESY data harder to analyze and NOESY spectra were generally used where possible in this thesis. For more on NOE/ROE theory, related to $\tau_c$, see Section 3.5.

### 2.1.3 NOE in structure determination

The NOE correlations are usually used qualitatively to determine stereochemistry or confirm the assignment of nuclei in novel or known compounds.[7,16] The information is highly local due to a $1/r^6$ dependence on the intensity of the observed signal and distances up to 4-5 Å are usually observed, depending on the compound size in the form of the rotational correlation time of the nuclei pair. The rotational correlation time is dependent on the size of the molecule, the temperature, the shape and flexibility of a structure and the viscosity of the solvent. The signal from NOE spectroscopy may be quantified if care is taken in the experimental setup and some approximations are made, presented below.

In general, two methods are available for obtaining quantitative distances from NOE data; with or without the use of a relaxation matrix.[17–21] The difference between these methods is in short the approach to handling multi-spin effects such as unwanted spin diffusion. Spin diffusion may be explained from a situation where three spins ($H_a$, $H_b$ and $H_c$) are considered, as in Figure 2.5.

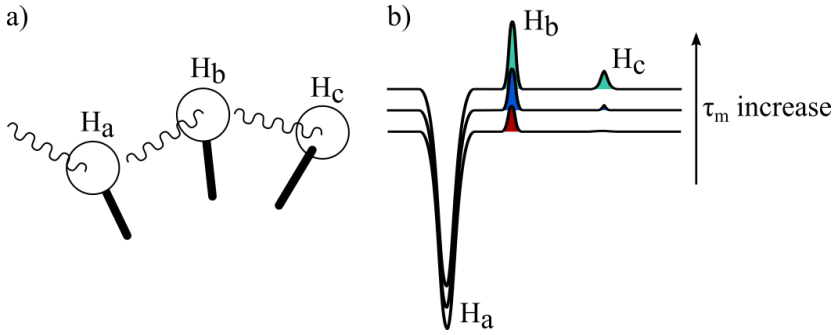



**Figure 2.5. Spin diffusion explained. a) Three spins are considered ($H_a$, $H_b$ and $H_c$). $H_b$ is positioned between $H_a$ and $H_c$, and is involved in NOE relaxation with both, while $H_a$ and $H_c$ are far from each other with negligible cross-relaxation. Magnetization is transferred from $H_a$ to $H_b$, and once sufficient magnetization has built up on $H_b$, magnetization transfers (or diffuse) from spin $H_b$ to $H_c$. b) An illustration of a 1D- or slice of a 2D NOESY spectrum shows the NOE cross-peak intensity increasing with an increase in the used mixing times.**

Spin diffusion will create a cross-signal between A and C with a stronger intensity than expected from the inter-proton distance, and the signal may lead to a false stereochemistry or experimental distance. This problem may be avoided or diminished by using short mixing times as magnetization has less time to build on B in Figure 2.5.[7,22] When determining whether to use a relaxation matrix method or not, it is important to consider the system; for large molecules where spin diffusion is rapid, it should be included in the calculations. For small molecules, using short mixing times, spin diffusion is less of a problem. Thus, using a matrix that also includes a multitude of approximations may not be needed or beneficial.

Another important consideration in the experimental setup is the utilized delay time between scans, which is true for all quantitative NMR. If inadequate $T_1$ relaxation is achieved for any protons during the delay, the intensity of subsequent scans will be altered accordingly leading to wrong relative intensities and thus wrong distances, see below.

### 2.1.4 Quantitative calculations

The nuclear Overhauser effect (NOE) gives through-space correlations from dipolar relaxation between nuclei. The relationship between cross-relaxation, $\sigma$, and distance, $r$, is given in equation (2.3) and rewritten below to match most litterature.[7]

$$\sigma_{IS} = \left(\frac{\mu_0}{4\pi}\right)^2 \frac{\hbar^2 {\gamma_I}^2 {\gamma_S}^2}{10} r^{-6} \left(\frac{6\tau_c}{1+4\omega^2{\tau_c}^2} - \tau_c\right) \quad \textbf{(2.3)}$$

This is easily simplified, assuming identical nuclei in all experiments, to equation (2.6), where $k$ covers the constants in (2.3) and is identical for all observed NOEs.

$$\sigma_{IS} = k\, r^{-6} \left(\frac{6\tau_c}{1+4\omega^2{\tau_c}^2} - \tau_c\right) \quad \textbf{(2.6)}$$

Assuming identical correlation time for all nuclei pairs, it is further simplified to equation (2.7). Again $k$ covers constants and is equal for all NOEs.

$$\sigma_{IS} = k\, r^{-6} \quad \textbf{(2.7)}$$

The cross-relaxation rate for transient NOEs is correlated to the observed intensity, $\eta$, by equation (2.8).

$$\eta_{IS} = \sigma_{IS}\tau_m \quad \textbf{(2.8)}$$

Where $\tau_m$ is the mixing time used in the experiment. If identical mixing times are used, the relationship of intensities and distances between two nuclei pairs may be defined as in equation (2.9).

$$\frac{\eta_1}{\eta_2} = \frac{k\, {r_1}^{-6}}{k\, {r_2}^{-6}} \rightarrow r_1 = r_2 \left(\frac{\eta_1}{\eta_2}\right)^{-\frac{1}{6}} \quad \textbf{(2.9)}$$

Thus, if a single distance is known, others may be calculated from their relative intensity. This relationship is known either as the isolated spin-pair approximation (ISPA) or initial rate approximation (IRA).[22,23] As indicated this is an approximation due to a couple of factors.

1. It is approximated that all rotational correlation times are equal, and unknown dynamics thus affect the NOE.
2. It is assumed that the acquired mixing time is short, eliminating any effect from spin diffusion.
3. The cross-relaxations of multiple involved spins are assumed to behave as isolated spins.

In order to avoid the inclusion of spin diffusion contribution to the NOE intensities in calculations, the mixing time may be varied to construct build-up curves to ensure that the utilized mixing time is in the linear regime, exemplified in Figure 2.6. For spin diffusion the onset of magnetization buildup will be later than for "true" NOE cross-peaks making them easily distinguishable. This also allows for the elimination of possible noise or artifact signals in the subsequent analysis, as these signals will be completely independent of the mixing time.

There are two methods to proceed once the linear range has been established. One is to use the cross-relaxation rate which may be determined directly from the lower graph of Figure 2.6 (right) as the intersection with the y-axis. The other is to utilize a fixed mixing time in the linear range. While the first method will effectively average out some of the internal intensity errors in the spectra, the latter is usually used in this thesis, due to interferences from COSY-type artifacts in the spectra, especially at low mixing times due to the lower NOE intensities.[7] To use points within the linear range, where the scalar coupling contribution is negligible, leads to the inclusion of more NOEs and thus more information and the associated errors were always evaluated to be within the general error of the experiments. It is here utilized that the scalar coupling intensities are less dependent on the mixing time than the NOE cross-peak intensities.

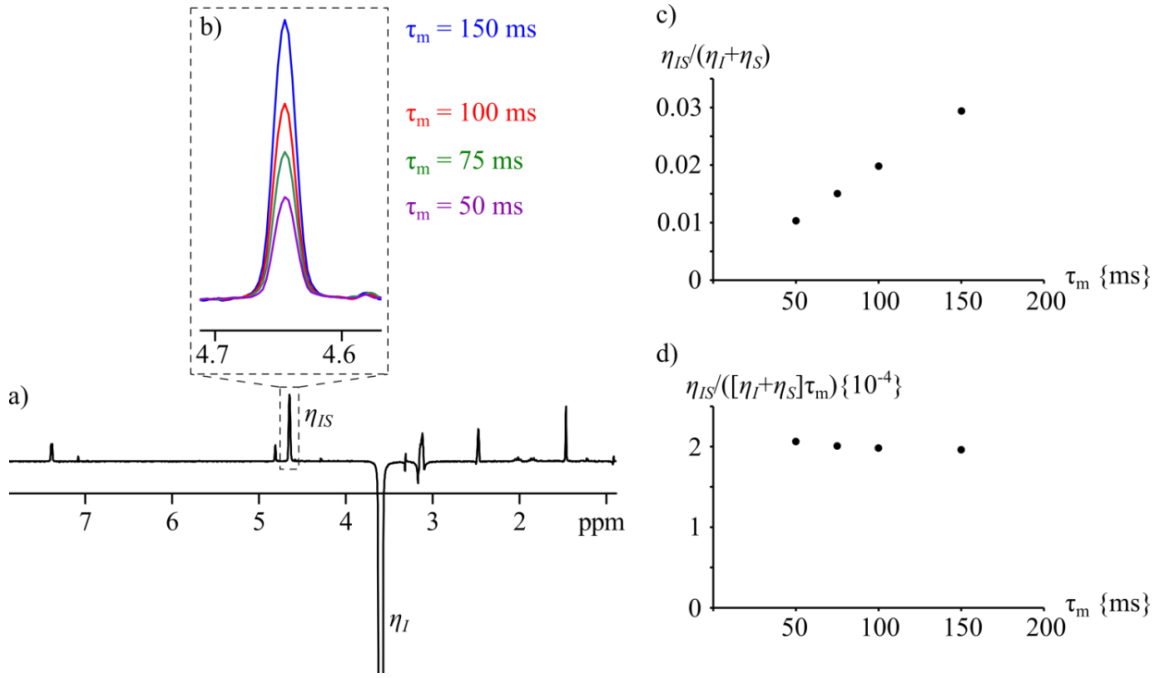



**Figure 2.6. a) 1D slide from 2D NOESY spectrum of epi-10,23-dihydro-24,25-dehydroaflavine (see Section 3.2.2). The dependence of the intensity on the mixing time ($\tau_m$) is seen in the enlargement (b) and is plotted on the right in a buildup curve for the exemplified NOE (c and d).**

In theory, any distance may be used as reference distance, as all distances are used relative to each other. In practice, the most common distances chosen are those known to be static; diastereotopic protons (~1.78 Å) or neighboring aromatic protons (~2.48 Å). The signal intensity of both of these may be modulated due to scalar couplings, which should be kept in mind. Furthermore, a dependence on the value of the chosen reference distance and the error distribution has been reported, as the distances of the same size resulted in a better fit between experimental and back-calculated data.[7] That being said, excellent fits have been reported utilizing e.g. diastereotopic protons, which are on the low side of relative distances in organic structures.[21] Note that the set distance of the reference may be allowed to differ (e.g. up to 5 % error), to obtain the best possible correlation between experimental and 3D structure distances. This is most easily achieved on rigid systems, with multiple fixed distances.

### 2.1.5 The 1D NOESY/ROESY (PANIC) approach

The PANIC or Peak Amplitude Normalization for Improved Cross-relaxation method is, other than a testament of the creative acronyms created by the NMR community, a straight forward implementation of the ISPA approach.[23] It is based on the work of Macura *et al.* discussed in the next section, but is used primarily for 1D NOESY/ROESY data.[17] It is used that the linear range of the NOE buildup curve is quite remarkably extended by implementing the auto-relaxation of the irradiated nuclei into the measurement of the NOE intensities in equation (2.10).[17,23]

$$\eta_{IS}{}^{PANIC} = \frac{\eta_{IS}}{\eta_I} \qquad \textbf{(2.10)}$$

In practice this is easily achieved by setting the integral of the irradiated peak to be equal among all 1D spectra obtained and these “PANIC intensities” are compared in the ISPA method. The resulting

relative intensities are thus corrected for auto-relaxation. The irradiated peak is set to a value of -1000, chosen for convenience as this makes the relative magnitude of the cross-peaks in the order of 0-500 for small molecules, dependent on $\tau_c$ and $\tau_m$.

### 2.1.6 The 2D NOESY/ROESY approach

As mentioned in the previous section, the PANIC approach is based on the work of Macura *et al.* for 2D spectra.[17,23] Thus, when dealing with 2D NOESY or ROESY spectra the methodology outlined here is used. It is essentially equal to the PANIC approach, but instead of using a single irradiated peak as reference, the cross-peaks ($\eta_{IS}$) are normalized by the average of the two appropriate diagonal peaks ($\eta_I$ and $\eta_S$) as in equation (2.11).

$$\eta_{IS}{}^{norm} = \frac{\eta_{IS}}{½(\eta_I + \eta_S)} \tag{2.11}$$

If a lack of resolution leads to overlap in the diagonal peaks, equation (2.12) may be used, as it is assumed that $\eta_I$ and $\eta_S$ are close to identical for equal mixing time and equal nuclei.

$$\eta_{IS}{}^{norm} = \frac{\eta_{IS}}{\eta_I} = \frac{\eta_{IS}}{\eta_S} \tag{2.12}$$

In practice it may be beneficial to divide overlapping diagonal peaks by the number of protons present, as the auto-relaxation is more likely to be equal for an identical electronic environment and use this average diagonal peak in the calculations.[24]

### 2.1.7 Relaxation matrix approach

The relaxation matrix approach was not generally utilized in the work described in this thesis, though it was evaluated initially. A short introduction will thus suffice. The Solomon equation, which is given in equation (2.1), may be re-written in matrix form where cross-peak intensities are considered as a function of mixing time, $\tau_m$.[7,22,25,26]

$$\frac{\partial A(\tau_m)}{\partial \tau_m} = -RA(\tau_m) \tag{2.13}$$

$$A(\tau_m) = A_0 e^{-R\tau_m} \tag{2.14}$$

Where $A(\tau_m)$ is the matrix of cross-peak intensities at $\tau_m$, and $R$ is the symmetrical relaxation rate matrix given in equation (2.15). $A_0$ is a diagonal matrix consisting of the cross-peak intensities at $\tau_m = 0$ s.

$$R = \begin{bmatrix} \rho_1 & \sigma_{12} & \cdots \\ \sigma_{21} & \rho_2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \tag{2.15}$$

The diagonal elements of the relaxation matrix are the dipolar longitudinal relaxation rate constants, $\rho$, and the off-diagonal elements are the cross-relaxation rate constants, $\sigma$. From these the inter-proton distance between nuclear spins may be obtained.

This may be re-arranged to (2.16) to express the relaxation rates in terms of NOE intensities.

$$-\ln\left(\frac{A}{A_0}\right)\tau_m = R \tag{2.16}$$

Which may be expressed and solved as (2.17), where the matrix *E* contains the eigenvector matrix and *N* contains the eigenvalues of $A/A_0$.

$$\frac{-E \ln N \; E^{-1}}{\tau_m} = R \tag{2.17}$$

Cross-relaxation rates can thus be derived and used to obtain inter-nuclear distances. The NOE intensities may also be back-calculated from a molecular 3D structure using the inverse of this method to generate a theoretical NOE spectrum, which can be iteratively compared to the experimental NOE spectrum for structure refinement and verification.[22,25,26]

If the matrix works perfectly in its implementation the method takes multi-spin effects, such as spin diffusion, into account. It is thus most relevant when dealing with macromolecular systems where multi-spin effects are generally more prevalent. Problems can occur for full relaxation matrix analysis where incomplete data sets (e.g. due to spectral overlap or spectral noise) make derivation of the full intensity matrix difficult. This is circumvented by combining experimental data with cross-peak intensities calculated from a theoretical model by programs that utilize the full relaxation matrix method such as CORMA (Complete Relaxation Matrix Analysis) and MARDIGRAS (Matrix Analysis of Relaxation for discerning the Geometry of an Aqueous Structure).[22,26]

### 2.1.8 Averaging of NOE data

When an average of different nuclei is observed, e.g. from overlapping resonances of methylene- or methyl groups, averaging methods are needed in order to cope. In non-methyl cases, equation (2.18) is used in distance calculations.

$$\eta_{I(Sa+Sb)} = k \, ({r_{ISa}}^{-6} + {r_{ISb}}^{-6}) \tag{2.18}$$

For methyl groups another approach is taken. The traditional methods to average methyl groups are derived for protein NMR and are split in two, dependent of the relative local rotational correlation time of the methyl group compared to the rest of the structure, given in equations (2.19) and (2.20).[27]

$$\eta_{IMe} = k \, \langle {r_{IMe}}^{-3} \rangle^2 \tag{2.19}$$

$$\eta_{IMe} = k \, \langle {r_{IMe}}^{-6} \rangle \tag{2.20}$$

If the methyl group is spinning much faster than the rotational correlation time for the general molecule, the NOE is averaged as (2.19) and if the effective rotational correlation times are close to equal (2.20) is used.[27] In practice (2.20) is only used for very small molecules in very low viscosity solvents.

### 2.1.9 Handling of rigid and flexible molecules

NOE distance analysis has proven very accurate for small molecules when few conformers are present.[21,28–30] From NOE distances another low abundance conformer was thus shown to be present for strychnine.[30] For more flexible compounds it is beneficial to rely on less optimized structures or on centroids which represent the full conformational space. The usage of centroids is a laborious task though; one may need to use many centroids if the compound is very flexible. One also has to identify precisely what parts are of interest for the structure as the nuclei used in establishing the

centroids. The inclusion of multiple structures requires means of averaging a vast amount of structures in a sensible way. Overfitting or fitting of data to a wrong conformational average may here present a problem since multiple different conformational populations may correlate well to the distances obtained from the experimentally observed NOEs.[31,32] It is thus recommended to couple the NOE data to other types of NMR data, being either *J*-coupling or residual dipolar coupling, see Sections 2.2 and 5.1. This will alleviate some problems due to the difference in the averaging mechanisms for the different data; averaging distances, dihedral angle, angles etc.

When comparing multiple structures to NOE derived distances, the average distances are calculated by equation

$$r_{ave} = \langle p_i r_i^{-6} \rangle^{-1/6} \quad \textbf{(2.21)}$$

Where $p_i$ is the population with a given distance. The averaging result in a larger influence of shorter distances to the observed average distance.[30]

In this thesis, multiple approached are used, each suited to the amount of flexibility expected, and indicated from simulations, for the given structure. For more information see the experimental section and the sections included in Chapter 3.

### 2.1.10 Error analysis

Though widely regarded a disadvantage or limitation, the distance to intensity relationship of $r^{-6}$ may actually be viewed as a clear advantage of the NOE over other 3D observables (*J*-coupling constants, RDCs etc.). This is due to the relationship between experimental error and resulting distance error, illustrated in Table 2.1.

**Table 2.1. Theoretical relationship between the error of the observed intensity and the error in the calculated distance both in %. Calculated using equation (2.3).**

| Observed $\eta$ error [%] | Resulting $r$ error [%] | Observed $\eta$ error [%] | Resulting $r$ error [%] |
|---|---|---|---|
| -10 | -1.8 | 10 | 1.6 |
| -20 | -3.8 | 20 | 3.0 |
| -30 | -6.1 | 30 | 4.3 |
| -40 | -8.9 | 40 | 5.5 |
| -50 | -12.2 | 50 | 6.5 |

This effectively means that rather large experimental errors, as a product of e.g. faulty assumptions of equal correlation times, spin diffusion or spectra with low *S*/*N*, will often result in quite small errors in the calculated distances, making the NOE approach highly favorable compared to other experimental methods. The difference on whether a too large or small NOE intensity is observed is due to the favoring of small distances from the $r^{-6}$ dependency.

The decrease in NOE intensities for longer distances will result in longer distances being more prone to errors, since the NOE intensity will approach the spectral noise and the experimentally obtained data will more easily differ considerably from the actual NOE.

## 2.2 *J*-coupling constants

The next method used to obtain 3D structural information is by indirect dipole-dipole couplings, in the form of scalar or *J*-couplings. The phenomenon was first reported as "low beats" in an echo from spin echo methods for ethanol by Hahn in 1950.[33] The origin of the observations was disputed until the Ramsey and Purcell formally introduced the *J*-coupling constant definition a couple of years later.[34–36]

Scalar couplings are magnetic interactions transmitted by the bonding electrons by which the spins are indirectly connected.[6,37] The coupling occurs as the magnetic moment of the nuclei polarizes electrons involved in the bond slightly, and this polarization is transmitted by overlapping orbitals to other nuclei. This leads to the spin state of one nucleus to influence the effective external magnetic field of neighboring nuclei and thus line splitting. The interaction is independent of the applied magnetic field since it only depends on the spin orientation and the orientation of the electrons which are paired due to the Pauli exclusion principle.[11]

The situation is depicted in Figure 2.7 over one bond for two bonded nuclei. The resulting signal will be a doublet due to the two different energy states, with a positive *J*-coupling constant.[11] The *J*-coupling constant is positive if the spin-spin coupling increases the energy when spin polarizations are parallel (triplet spin state) and decreases the energy when the spin polarizations are anti-parallel (singlet spin state).[11,36] Negative *J*-coupling constants have an opposite dependence. This assumes an equal sign of the gyromagnetic ratio of the two nuclei.[11]

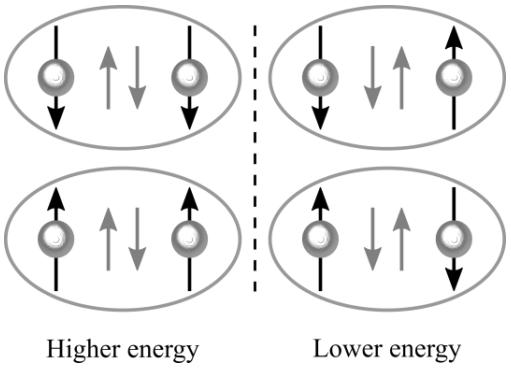



**Figure 2.7. Mechanism of scalar coupling between two neighboring spins (over one bond). The black arrows are the spin angular momentums and the grey arrows are the electron spins, paired due to the Pauli principle. The nuclear spins affect the spin of the electrons, and opposite polarizations of the nuclear spins are favored energetically as the electron spin polarizations are kept opposite. An equal nuclear spin polarization is high in energy as no favorable spin orientation may be achieved. Inspired by Levitt.[11]**

The energy of the spin-spin interaction, given as the Hamiltonian, depends on the magnetic moments and the position of nuclear spin vectors (*I*) of the involved nuclei as given in equation (2.22).[6,11,38]

$$\hat{\mathcal{H}}_{IS}^{J} = J_{IS}\hat{\mathrm{I}}_I \cdot \hat{\mathrm{I}}_S \qquad \textbf{(2.22)}$$

Implicit in the *J*-coupling constant is the product of the magnetogyric ratios of the involved nuclei and the reduced constant *K* is sometimes used, where this dependence is eliminated.[6] *K* is not used in this thesis.

Ramsey identified three interactions, which contribute to the *J*-coupling with the first being the most dominant:[36]

1. Fermi contact (FC) interaction of electronic and nuclear spins.
2. The electron orbital motion with nuclear spin.
3. Dipolar terms involving electronic and nuclear spins.

For the FC term which may be seen as a direct contact term of electrons and nuclei, σ electrons play a significant role in transmission while π electrons are mostly involved due to exchange interactions between the σ- and π-electronic systems.[39]

The scalar couplings lead to the observed multiplets in e.g. 1D $^{1}$H spectra and may be used in structure elucidation. Scalar couplings are also the foundation of most NMR experiments which employ coherent magnetization transfers e.g. COSY (mostly $^{2/3}J_{HH}$), HSQC ($^{1}J_{CH}$) or HMBC ($^{n}J_{CH}$) type experiments.[6]

### 2.2.1 $^{3}J_{HH}$-coupling constants

The $J_{HH}$-coupling constants in the form of $^{n}J_{HH}$ ($n$=2,3), are arguably the easiest 3D structural parameter to extract as they are present in simple 1D $^{1}$H spectra, and may be extracted when the resonances are not overlapping. If other types of coupling constants or a higher resolution, due to congested 1D spectra, are wanted, tailored experiments are usually needed; see Chapter 4 for a more elaborate discussion. The size of the *J*-coupling constants is, as mentioned above, based on the distance and degree of orbital overlap between the relevant nuclei. The influence of orbital overlap is the basis for the sinusoidal relationship between the size of coupling constants and a dihedral angle, which for $^{3}J_{HH}$-coupling constants are given by the semi-empirical Karplus or Karplus-like equations as illustrated in Figure 2.8.[6,40]

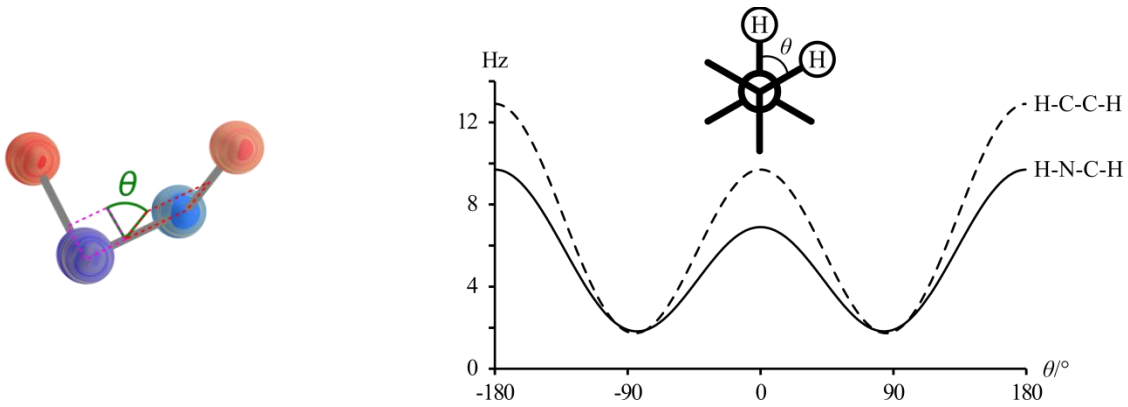



**Figure 2.8. Left: Dihedral angle between protons (red) three bonds apart, blue may be carbons or heteroatoms. Right: Dependence of $^{3}J$-coupling constant on dihedral angle for peptide data as given by Karplus equations.[40]**

The constants present in the Karplus equations have generally been determined empirically from experimental data, though computational data may also be used, if e.g. less experimental data is present or to fully cover the given Karplus curve.[40] The relative size of the *J*-coupling constants is also dependent on the nuclei involved, e.g. HN-CH versus HC-CH coupling constants due to the difference in magnetic moment, and the properties of surrounding nuclei. This dependence is utilized in e.g. the Haasnoot, De Leeuw and Altona (HLA) or Altona equation.[40–42] Also the H-C-C

internal angle, the C-C and N-H bond length and structural strain have been shown to influence the size of the coupling constants.[11,36,43–45] The Karplus equations generally take the form of equation (2.23).[40]

$$^{3}J_{\mathrm{HH}}(\theta) = A\cos^2\theta + B\cos\theta + C \tag{2.23}$$

Where $\theta$ is the dihedral angle and the constants *A*, *B* and *C* depend on the type of coupling as addressed above.[40] The equation was derived based on peptide moieties, and the amide version was used for NH-CH coupling constant calculations where $A$ = 6.4, $B$ = -1.4 and $C$ = 1.9.

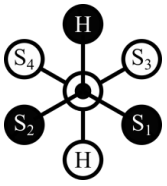
An example of a more elaborate equation is the HLA equation given in equation (2.24), which was established to take the electronegativity of the neighboring nuclei into account.[42]

$$^{3}J_{\mathrm{HH}}(\theta) = P_1\cos^2\theta + P_2\cos\theta + P_3 + \sum \lambda_i(P_4 + P_5\cos^2(\varepsilon_i\theta + P_6|\lambda_i|))$$
$$\lambda_i = (X_\alpha - X_H) + P_7\sum(X_\beta - X_H) \tag{2.24}$$

This is the equation used for $CH_n$-$CH_m$ coupling constants in this thesis unless stated otherwise, used with the appropriate constants from Table 2.2.

**Table 2.2. Left: Constants for use in HLA calculations for different substitution patterns of the carbon atoms. Right: Relative orientation of substituents included in the HLA calculations.**

| Type | $P_1$ | $P_2$ | $P_3$ | $P_4$ | $P_5$ | $P_6$ | $P_7$ |
|---|---|---|---|---|---|---|---|
| $CH_3$-$CH_2R$, $CH_3$-$CHR_2$, $RCH_2$-$CH_2R$ | 13.70 | -0.73 | 0.00 | 0.56 | -2.47 | 16.90 | 0.14 |
| $R_2CH$-$CH_2R$ | 13.22 | -0.99 | 0.00 | 0.87 | -2.46 | 19.90 | 0.00 |
| $R_2CH$-$CHR_2$ | 13.24 | -0.91 | 0.00 | 0.53 | -2.41 | 15.50 | 0.19 |




The post $P_3$ term is the sum of the attached groups and is dependent on the electronegativity ($X$) of the substituents ($S_{1\text{-}4}$), see Table 2.2 for clarifications. $\varepsilon_i$ is equal to 1 for $S_1$ and $S_3$, and -1 for $S_2$ and $S_4$.

Unlike the more complicated averaging of NOEs presented in the previous section, *J*-coupling constants are usually averaged over multiple structures as the weighted average of the *J*-coupling constants for the different conformers. It is thus not averaged as the weighted average of the dihedral angle, an important distinction to make. The averaging of *J*-coupling constants may thus be considered orthogonal to the averaged NOEs.

## 2.3 Evaluating data

To evaluate e.g. back-calculated distances from 3D structures to the relative distances obtained from NOEs, the mean absolute error (MAE) is used directly or on the percentage error, equations (2.25) and (2.26). The MAE is also used for *J*-coupling constants.

$$\mathrm{MAE} = \frac{1}{n}\sum\left|x_i^{exp} - x_i^{calc}\right| \tag{2.25}$$

$$\mathrm{MAE}(\%) = \frac{1}{n}\sum\frac{\left|x_i^{exp} - x_i^{calc}\right|}{x_i^{exp}} \tag{2.26}$$

## 2.4 Introduction to computational chemistry

Often 3D structures are desired for discussing observation such as NMR data, biological assays or chemical properties in structural terms.[46–48] For the generation of optimized structures a plurality of methods are available.[49] In the literature multiple approaches have been proposed which may be used for structural calculation, and since the rationale behind the approaches taken in this thesis is discussed elsewhere in the text, only a short introduction is included here.[49,50] Also, the theories, quantum chemistry and formula which govern the methods are only very briefly touched upon.

### 2.4.1 Molecular mechanics and dynamics

Molecular mechanics (MM) and molecular dynamics (MD) is a force field (FF) method, where the chemical bonds are described by classical mechanics where the positions of the nuclei are considered. MM is used in calculating the energy of a given system and e.g. differences in lengths ($E_{str}$), angles ($E_{bend}$) and dihedral angles ($E_{tors}$) between structures are scored in order to identify the lowest energy structure(s).[49] Non-bonded interactions such as electrostatic ($E_{el}$) and van der Waals ($E_{vdw}$) interactions are also included. The energy of the structure will thus increase by deviations in distances, angles etc., and the overall energy of the system will be determined by an energy calculation as the one given in (2.27).[49]

$$E_{FF} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{el} + E_{cross} \quad \textbf{(2.27)}$$

The term $E_{cross}$ essentially covers combinations of the bonded interactions to further improve the force field when matched to experimental data. For example, the force field used in this thesis for organic solvents, MMFF, uses a combination of stretch and bend interactions for this term.[49] In the work of this thesis, the MM calculations were always coupled to conformational sampling, e.g. by Monte Carlo methods where the geometry is perturbed by varying e.g. torsion angles at random before minimization, or MD, see below.[49,50] This results in optimized structures which generally should be good approximations to actual minima, but as each force field may score parameters differently, minima will be force-field dependent. It is thus not assumed that the relative energy between different compounds or conformers may be taken as accurate measures, and the structures are always compared to experimental data. It is thus important that the used methods result in structures of the global as well as local minima, and that the conformational space is thoroughly sampled.

Also MD was used for simulations. In short, MD simulations treat the molecules as classical Newtonian systems, where the different interactions in between nuclei, such as bonds, produce a force on the nuclei. The system is then evolved for a given time, and the equations are continuously calculated in given time-steps.[49,50] Since the time-steps are often short a great dependence on starting structure is often found for the simulations, and simulated annealing, where the temperature is varied (usually from high to low) may help this, as the structures may traverse larger energy barriers.[49]

The solvents are included in the simulations, since the experimental properties observed by NMR in this thesis are observed in the liquid state. The treatment of solvents in MM calculations, as well as Density Functional Theory (DFT) calculations is complex, and will not be discussed. Briefly, the solvent implementation may either implicit, where the solvent is treated as a continuous medium,

usually where the properties of the solvent is given by a dielectric constant that interact with the molecule, or explicit where multiple solvent molecules are included in MD simulations and evolved with the molecule of interest.[49,50]

### 2.4.2 Density functional theory

Density Functional Theory (DFT) is a computational approach for calculation of electron energies. Here, approximations are needed as the electron energy cannot be determined or estimated by classical mechanics, and a set of functions is used to determine the energy of a system by the electron density of said system.[49,50]

The basis of DFT calculations is formed by the functionals which are the series of formulae used to describe and solve the electron density. Multiple functionals have been derived and are available, often from a combination of different exchange and correlation functionals.[49] The DFT functionals used in this study are the Becke Three Parameter Hybrid Functionals (B3LYP), probably the most widely used method for small molecules, and MPW1PW91, which is a single parameter functional developed in part for NMR calculations.[49,51] Both are hybrid functionals and include a mixture of Hartree-Fock exchange and DFT exchange-correlation.[49]

The functionals use basis sets which may also be varied according to the investigated problem. A basis set consists of predetermined functions which are used to describe molecular orbitals centered at each nuclei using a linear combination of Gaussian type orbitals (GTOs).[49] The GTOs are used to approximate Slater-type orbitals (STOs) by linear combinations. STOs are not determined directly due to computational inefficiency. The basis set may be augmented by polarization functions or diffuse functions. Polarization functions add another orbital to the initially used orbitals for a given nucleus. For a nucleus with and outer *p*-orbital a *d*-orbital may thus be added, which allows of more asymmetry around the nucleus. Diffuse functions are Gaussian functions added to better describe the electron density far from the nucleus.[49]

When applying DFT calculations to structural challenges the Pople basis set was used throughout this thesis. These basis sets are built as exemplified for 6-31g(d); the core orbitals are described by six GTOs, the inner part of the valence orbitals are described by three GTOs and the outer part of the valence orbitals by one GTO. The (d) means that a polarization function is used and a *d*-type polarization function is added to the heavy atoms. Other type of basis sets could be the correlation consistent basis sets e.g. cc-pVDZ. The correlation consistent basis sets differ from the Pople set as they are designed to converge towards the basis set limit, where properties are described as if using an infinite basis. As such a greater basis set should always lead to better calculations of the property of interest. It was recently reported that the basis set limit, where the calculated properties converge, was reached for the calculation of NMR properties, which should lead to an even stronger coupling of experimental NMR spectroscopy and computational chemistry.[52]

It is important to choose a basis set that serves the need of the system in question. Larger basis sets may lead to increase in accuracy if the orbitals are better described but will also increase the computational time. The increase in accuracy is not always realized though as an observable may be better parameterized by a smaller basis set when using the Pople basis sets. It is thus often useful to screen different DFT functionals and basis sets in order to obtain a match for a given challenge, or

to take inspiration for methods which have been shown to work well for small molecules by others.[47,53,54] For the optimization of structures it is always useful to slowly increase the basis set as the computational time increases with size starting from e.g. output from FF methods. The computation time may thus be minimized by submitting already optimized structures to the more expensive calculations. After optimization the resulting structures were used for the calculations of NMR observable properties, namely chemical shifts and *J*-coupling constants. Whether it is needed to optimize structures to a DFT-level of theory for chemical shift calculations has been questioned, but it seems to be needed for calculations of *J*-coupling constants.[54–58]

### 2.4.3 Chemical shifts

The chemical shift is defined as the difference in shielding tensors of nuclei compared to the shielding tensor of a reference, which for $^{1}$H and $^{13}$C nuclei are the shielding tensors of tetramethylsilane (TMS). Similarly the calculated property from DFT calculations is the NMR shielding tensor, which is a second order magnetic property, defined as the energy derivate of the nuclear magnetic moment $\mu_I$ with respect to the applied magnetic field, *B*.[52]

$$\sigma_I = \partial^2 E / \partial \mu_I \partial B|_{B,M_I=0} \quad \textbf{(2.28)}$$

The calculations of shielding tensors are based on the interaction of the applied magnetic field, the magnetic field generated by the nuclei and the magnetic field generated by the rapid movement of the electrons in a molecule. [47,54,58,59]

The first two terms are termed static fields and perturb the kinetic energy of the Hamiltonian so that if a finite basis set is used to model the electron distribution, the operator depends on the origin of the coordinate system known as the gauge. This is circumvented by either the GIAO (gauge invariant atomic orbitals) method, which uses basis functions that have an explicit field dependence, or the CSGT (continuous set of gauge transformation) which performs a gauge transformation at each point in space. It has been shown that for the nuclei involved in NMR the two methods work equally well, but that the GAIO method converges faster.[47]

Generally two methods are available for correlating experimental chemical shifts to the calculated GIAO nuclear magnetic shielding tensors. One is to calculate the tensors for TMS and use these tensors as reference in comparison to an internal standard for liquid NMR. There is a possible caveat as the GIAO tensors of TMS must be calculated for each functional and basis set used, and the silicon in TMS is not necessarily parameterized well in all of the used basis sets, a problem especially for smaller basis sets.[47] The second, and more general, approach is to linearly correlate the shielding tensors to the observed chemical shifts, as in (2.29).[54,58]

$$\delta_{scaled} = \frac{\sigma_{GIAO} - b}{a} \quad \textbf{(2.29)}$$

Where *a* is the slope and *b* the intercept of the correlation of observed shifts and calculated shielding tensor. The approach is simpler and less computationally demanding but could potentially result in overfitting of the data, as each structure is individually scaled to experimental data. The approach is used in this thesis in line with most literature.[47,54,58]

### 2.4.4 *J*-coupling constants

Nuclear spin-spin coupling constants between two nuclei are calculated by the second derivatives of the energy with regard to their magnetic moments as seen in equation (2.30).[39,53]

$$J_{IS,xy} = \partial^2 E / \partial \mu_{I,x} \partial \mu_{S,y} \quad \textbf{(2.30)}$$

As mentioned in an earlier section the coupling of the nuclei is comprised of the sum of contributions of different terms covering different mechanisms: The spin hyperfine interaction and the spin-orbit (SO) coupling.[39,53] The SO interactions may be divided into a diamagnetic (DSO) and a paramagnetic (PSO) component and depend on the magnetic moments created by the movement of electrons. The hyperfine interactions are comprised of the Fermi contact (FC) term and the spin-dipole (SD) operator and depend on the interaction of an electron's spin with the local magnetic field of the nuclei.[53] The FC term describes the interaction at the nucleus and is usually, dependent of the nature and environment of the nuclei, by far the largest contribution to the total coupling constant. For proton-proton coupling constants all other terms than the FC may be neglected, and the result is scaled to correct for this and other possible inherent errors, an approach which has been shown to work well for small molecules.[53] The coupling constants are returned in Hz and are thus immediately comparable to the experimental data. Unlike shielding, scalar coupling calculations will be very sensitive to the shape of the electronic wave-functions near the nucleus, which should be modelled as accurately as possible for the FC term. The calculations are thus more demanding than the chemical shift calculations, as this is hard to simulate using GTOs.[39] The methods used for the calculations of *J*-coupling constants are discussed in the text in Chapter 4.

# 3 Application of NOEs and $^3J_{HH}$-couplings in 3D structure determination

## 3.1 Natural products

*The experimental work of this section is based on the articles below. The structures of homomorphosin A-F were elucidated during work done on the Master's Thesis "Chemistry of Black Aspergilli" by the author. A description of the purification and elucidation of the compounds is thus not included here, and relevant information may be found in the articles given below. The elucidation of the 3D structural features of homomorphosin A and homomorphosterol was achieved during the timeframe of this thesis. The structural elucidation and 3D structural calculations of the cyclomorhosins were done entirely during the timeframe of this thesis.*

Appendix A11: Dereplication Guided Discovery of Secondary Metabolites of Mixed Biosynthetic Origin from Aspergillus aculeatus
Appendix A11: Homomorphosins A-F, novel diketopiperazines

### 3.1.1 Introduction to natural products

Natural products are the collected term of metabolites from natural sources, usually microbes (fungi or bacteria) or plants. Metabolites may be primary, those crucial for the growth of the organism and thus shared by many species, or secondary metabolites (SMs), where the latter are generally small to medium sized molecules with another purpose than growth.[60,61] The functions of the SMs are not always known, but e.g. defence mechanisms, attractants or colouring agents have been suggested.[62,63] The discovery of penicillin in 1929 spurred the interest of in SMs due to their applicability as drugs and still the majority of approved drugs are of biological origin or derivatives of natural products.[64,65]

The SMs are divided into classes as polyketides, terpenes, non-ribosomal peptides and alkaloids, based on the origin of the precursors for the SMs and the genes utilized in the biosynthesis.[60] The biosynthesis is catalysed by enzymes such as polyketide synthases and non-ribosomal peptide synthetases, where covalent acyl-enzyme intermediates are formed and followed by condensations and tailoring to give the desired product, which are often enantiomerically pure compounds with many stereocenters and complex ring systems.[61,66,67] The molecules, and derivatives thereof, have usually been synthesised by total synthesis for further investigations, but the emerge of synthetic biology, where the machinery of the microbes are used to biosynthesise the wanted compound is a possible future in drug discovery.[68]

### 3.1.2 Aspergillus homomorphus

*Aspergillus homomorphus* (*A. homomorphus*) is a fungus residing in the *Aspergillus* section *Nigri* (black aspergilli), seen in Figure 3.1.[69,70] It is geographically highly specific and has only been identified from soil in the Dead Sea area and from an Argentinian vineyard.[71,72] It is thus a relatively rare fungus and the metabolic profile was previously only sparsely described.[73]

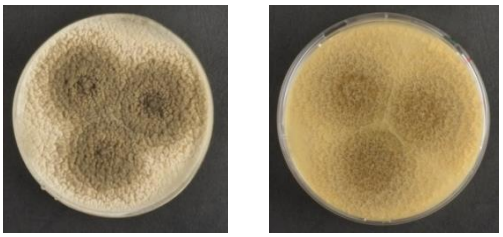


**Figure 3.1. *A. homomorphus* fungi grown for 7 days on YES media, left is top- and right bottom view.**

### 3.1.3 Homomorphosin A-F

The homomorphosins A-F are depicted in Figure 3.2. They are all diketopiperazines made from tryptophan and either valine, 2-aminobutyrate or alanine, and attached isoprene units. For the elucidation of the structures, the reader is referred to Appendix A11.

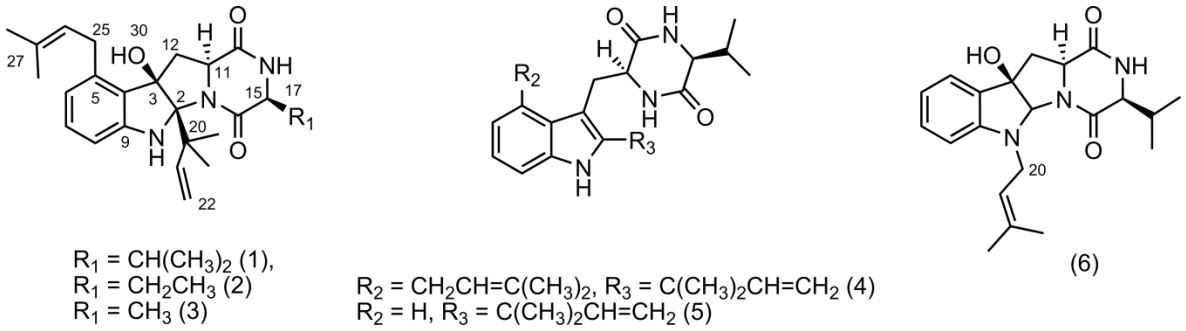



**Figure 3.2. Novel NRPs from *A. homomorphus*; Homomorphosin A-F (1-6). Homomorphosin A (1) is the main secondary metabolite produced by *A. homomorphus*.**

Homomorphosin A is the most predominant metabolite produced by *A. homomorphus* on any growth medium, and was subjected to a 3D structural analysis to verify the structure and to obtain structural insights. The calculations may also serve as an introduction to the techniques utilized throughout this chapter, from the simplicity of the structure. Distances from NOEs were obtained as described in the theory, by relative intensities in the linear range. The solvent was DMSO-$d_6$.[I]

Using a single structure approach, the best fit of the experimental data and a 3D structure is seen in Table 3.1 for NOEs and in Table 3.2 for *J*-coupling constants. Major violations in the NOE derived distances were observed for the flexible parts of the molecule which needed to be averaged.

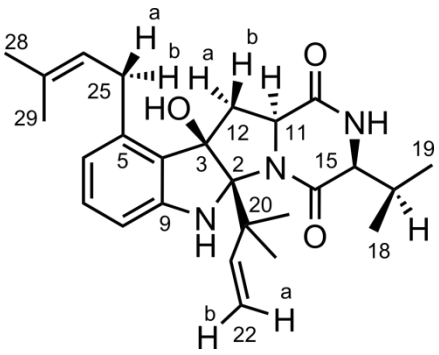
[I] Note that the solvents when acquiring NMR experiments throughout this thesis are always deuterated, unless explicitly stated otherwise.

**Table 3.1. Comparison of the distances from NOE intensities and 3D structures using a single structure, with the lowest MAE, optimized to B3LYP/6-31g(d) ($\tau_m$ = 200 ms). All distances in Å. Lower and upper bounds are defined as ± 10 % of the distances from the 3D structures. Rigid distances are in bold. [a]Reference distance used.**

| H1 | H2 | Exp. dist. | Lower bound | Violation | Upper bound | Violation |
|---|---|---|---|---|---|---|
| **1** | **8** | **2.66** | **2.39** | | **2.92** | |
| 1 | 21 | 2.88 | 4.18 | 1.30 | 5.11 | |
| 1 | 23/24 | 2.44 | 2.10 | | 2.56 | |
| 6 | 25a | 2.78 | 2.70 | | 3.30 | |
| 6 | 26 | 2.92 | 2.71 | | 3.31 | |
| 6 | 29 | 3.67 | 3.32 | | 4.05 | |
| **7** | **8** | **2.48[a]** | **2.23** | | **2.73** | |
| **11** | **15** | **2.46** | **2.20** | | **2.69** | |
| **14** | **15** | **2.72** | **2.46** | | **3.00** | |
| 14 | 17 | 2.94 | 3.33 | 0.39 | 4.07 | |
| 15 | 18 | 3.05 | 3.17 | 0.12 | 3.88 | |
| 15 | 19 | 2.58 | 2.25 | | 2.75 | |
| 15 | 17 | 2.53 | 2.27 | | 2.77 | |
| **17** | **18** | **2.30** | **2.08** | | **2.54** | |
| **17** | **19** | **2.30** | **2.07** | | **2.53** | |
| 21 | 23/24 | 2.47 | 2.26 | | 2.77 | |
| 21 | 12a | 2.87 | 2.44 | | 2.98 | |
| 21 | 30 | 2.80 | 2.96 | 0.16 | 3.62 | |
| 22b | 23/24 | 2.28 | 2.07 | | 2.53 | |
| 22b | 18 | 4.30 | 3.69 | | 4.51 | |
| 25b | 29 | 2.59 | 2.46 | | 3.00 | |
| 25b | 12b | 2.26 | 1.88 | | 2.30 | |
| **26** | **28** | **2.39** | **2.15** | | **2.63** | |
| 30 | 23/24 | 2.41 | 2.38 | | 2.91 | |
| 30 | 12a | 2.84 | 1.92 | | 2.35 | 0.49 |
| 30 | 25b | 2.95 | 2.89 | | 3.53 | |

**Table 3.2. Comparison of experimental *J*-coupling constants and calculated constants by the HLA equation from 3D structures using best fit single structure, in Hz.**

| H1 | H2 | *J* exp. | *J* calc. | Diff |
|---|---|---|---|---|
| 11 | 12b | 7.3 | 8.0 | 0.7 |
| 15 | 17 | 2.0 | 2.3 | 0.3 |
| 11 | 12a | 11.5 | 10.3 | 1.2 |




If multiple structures were used the correlation between the NMR and structural data was much better, and it was verified that the observed differences in distances for a single structure originated from structural flexibility, see Table 3.3, Table 3.4 and Figure 3.3. The utilization of multiple structures allowed the assignment of the "pro-chiral" methyl groups 18 and 19 as well as the diastereotopic protons 25 from an increase in MAE upon inversion, see Appendix A4. A constraint in the rotational space of the valine side-chain was initially suggested due to the low $H_\alpha$-$H_\beta$ coupling constant and the difference in the methyl chemical shifts and which was confirmed from the NOE distances. The methyl groups 23/24 were not solved due to similar and slightly overlapping intensities, and they are treated as the summed intensities in the calculations.

**Table 3.3. Comparison of the distances from NOE intensities and 3D structures using multiple structures from MM ($\tau_m$ = 200 ms). All distances in Å. [a]Reference.**

| H1 | H2 | Exp. dist. | 3D Dist. | Diff (Å) | Diff (%) |
|---|---|---|---|---|---|
| **1** | **8** | **2.61** | **2.66** | **0.05** | **1.9** |
| 1 | 21 | 2.88 | 2.96 | 0.08 | 2.7 |
| 1 | 23/24 | 3.34 | 3.37 | 0.03 | 0.9 |
| 6 | 25a | 2.78 | 2.76 | 0.03 | 0.9 |
| 6 | 26 | 2.79 | 2.80 | 0.01 | 0.3 |
| 6 | 29 | 4.41 | 4.42 | 0.01 | 0.1 |
| **7** | **8** | **2.48**[a] | **2.48** | **0.00** | **0.0** |
| **11** | **15** | **2.53** | **2.44** | **0.09** | **3.7** |
| **14** | **15** | **2.72** | **2.73** | **0.00** | **0.1** |
| 14 | 17 | 2.94 | 2.93 | 0.01 | 0.5 |
| 15 | 18 | 3.42 | 3.50 | 0.08 | 2.3 |
| 15 | 17 | 2.25 | 2.52 | 0.27 | 12.1 |
| 15 | 19 | 2.76 | 3.01 | 0.25 | 9.1 |
| **17** | **18** | **2.81** | **2.62** | **0.20** | **7.0** |
| **17** | **19** | **2.75** | **2.61** | **0.14** | **5.0** |
| 21 | 12a | 2.87 | 2.96 | 0.09 | 3.1 |
| 21 | 23/24 | 3.10 | 3.28 | 0.18 | 5.8 |
| 21 | 30 | 2.77 | 2.89 | 0.12 | 4.2 |
| 22b | 18 | 4.71 | 4.67 | 0.04 | 0.8 |
| 22b | 23/24 | 3.05 | 3.16 | 0.11 | 3.6 |
| 25b | 12b | 2.18 | 2.27 | 0.09 | 4.1 |
| 25b | 29 | 3.11 | 3.07 | 0.04 | 1.3 |
| **26** | **28** | **2.96** | **2.82** | **0.14** | **4.6** |
| 30 | 12a | 2.84 | 2.75 | 0.09 | 3.1 |
| 30 | 23/24 | 3.09 | 3.35 | 0.26 | 8.5 |
| 30 | 25b | 2.81 | 2.92 | 0.11 | 4.0 |
| | | | MAE | 0.10 | 3.5 |

**Table 3.4. Comparison of experimental *J*-coupling constants and calculated constants by the HLA equation from 3D structures using multiple structures from MM, in Hz.**

| H1 | H2 | *J* exp. | *J* calc. | Diff |
|---|---|---|---|---|
| 11 | 12b | 7.3 | 6.5 | 0.8 |
| 11 | 12a | 11.5 | 10.2 | 1.3 |
| 15 | 17 | 2.0 | 2.0 | 0.0 |

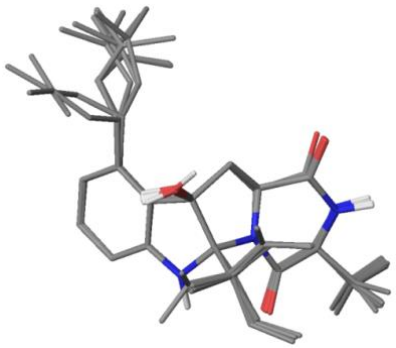


**Figure 3.3. Representation of the structural space inhabited by homomorphosin A, suggested from NOE and *J*-coupling constant data. The heavy atoms in the tetracyclic system of the 10 most highly populated conformers are overlain. Only the hydrogens of polar groups are shown.**

In conclusion the 3D structural investigation confirmed the structure obtained from other NMR experiments including the usage of qualitative NOEs. This is not surprising due to the rigid structure, but still the ability to determine the distances reliably dismisses any doubts. Also the accuracy is a good example of the degree of correlation between experimental and theoretical data that may be envisioned for the following small molecules, including the fact that rotatable bonds will be harder to fit than rigid parts, which is not surprising. Still, a pretty good fit of the experimental data was obtained from groups that exhibit freely rotatable bonds, and it is shown that allowing multiple conformers lead to a better fit of these parts, enabling supported suggestions to be made regarding stereochemistry. The ability to average over multiple conformers will be needed in the following sections.

The absolute stereochemistry was solved by Marfey's reaction as the valine was L-valine. Thus all stereocenters could be solved in relation.

### 3.1.4

### 3.1.5 Cyclomorphosins

Two compounds were isolated with the molecular masses $[M+H]^+$ of 850.3307 and 866.3611 m/z, determined from HRMS analyses. This translated into multiple possible molecular formulae for each structure, as the number of possibilities increase with size due to an increase in possible nuclei and associated isotope patterns. The most probable constituent formulae were $C_{46}H_{43}N_9O_8$ or $C_{45}H_{47}N_5O_{12}$ and $C_{47}H_{47}N_9O_8$ or $C_{46}H_{51}N_5O_{12}$, respectively for the two masses. In each case the latter constituent formulae were identified as the correct ones from NMR spectroscopy.

**Structural elucidation**

The structural elucidation process is exemplified for cyclomorphosin A (CM-A), and a similar approach was used to elucidate the structure of B, but from a significantly lower amount of compound. The structures of both NRPs are shown in Figure 3.4.

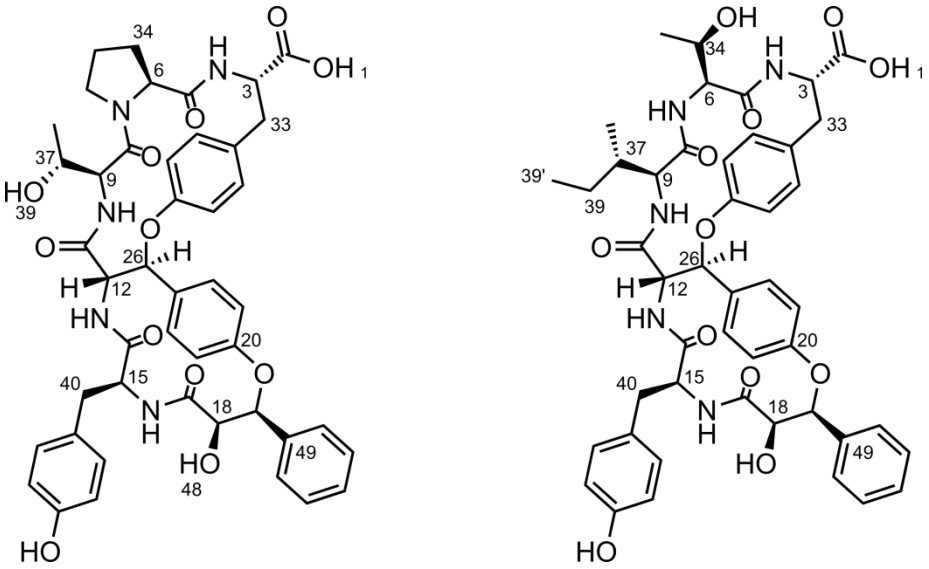


**Figure 3.4. Novel NRPs from *A. homomorphus*: Cyclomorphosin A (left) and B (right).**

The 1D $^1$H spectrum of CM-A is found in Figure 3.5. The difficulty of elucidation was increased by the presence of a major and a minor conformer in the ratio 3:1 as illustrated in the inset of Figure 3.5. For CM-B the ratio was determined to be approximately 10:1. The appearance of the resonances is the result of a second conformational species being present which interchanges slowly with the primary conformational species on the NMR chemical shift time scale. Luckily, only few resonances overlap between the conformers making it possible to unambiguously assign the shifts to each conformer.

**Table 3.5. Chemical shifts and HMBC correlations, assigned for CM-A and CM-B in DMSO-$d_6$.**

| | CM-A | | | CM-B | | |
|---|---|---|---|---|---|---|
| # | $\delta_H$ [ppm] (Int, mult, *J* [Hz]) | $\delta_C$ [ppm] | HMBC | $\delta_H$ [ppm] (Int, mult, *J* [Hz]) | $\delta_C$ [ppm] | HMBC |
| 1 | 12.76 (1H,s) | - | | ND | - | |
| 2 | - | 172.6 | | - | ND | |
| 3 | 4.60 (1H,ddd,11.5,8.5,4.7) | 51.0 | 2,33 | 4.50 (1H,td,10.1,3.6) | 52.2 | |
| 4 | 8.03 (1H,d,8.7) | - | 5,33 | 7.83 (1H,d,9.4) | - | |
| 5 | - | 170.6 | | - | 168.3 | |
| 6 | 4.75 (1H,d,6.7) | 58.9 | 34,35,36 | 4.19 (1H,dd,8.0,6.2) | 57.5 | 5,8,34,35 |
| 7 | - | - | | 6.87 (1H,d,6.8) | - | |
| 8 | - | 165.8 | | - | 170.0 | |
| 9 | 3.73 (1H,t,4.4) | 55.9 | 8,37,38 | 4.13 (1H,t,8.7) | 58.1 | 8,37,39 |
| 10 | 7.60 (1H,d,4.6) | - | 8,9,11 | 8.53 (1H,d,9.4) | - | 11 |
| 11 | - | 168.2 | | - | 169.2 | |
| 12 | 4.90 (1H,t,9.6) | 58.0 | 11,14,26 | 4.66 (1H,dd,9.7,7.3) | 58.6 | 14,26 |
| 13 | 8.21 (1H,d,10.2) | - | 14 | 8.33 (1H,d,9.9) | - | 14 |
| 14 | - | 169.4 | | - | 169.3 | |
| 15 | 4.14 (1H,dt,9.2,7.9) | 52.5 | 14,40 | 4.39 (1H,dt,9.3,8.0) | 51.7 | 17 |
| 16 | 6.56 (1H,d,9.1) | - | 15,17 | 6.48 (1H,d,9.5) | - | |
| 17 | - | 169.9 | | - | 169.1 | |
| 18 | 3.75 (1H,d,5.5) | 74.1 | 17,49 | 3.75 (1H,mult) | 74.2 | |
| 19 | 5.95 (1H,s) | 78.2 | 17,18,20,49,52 | 5.93 (1H,s) | 78.7 | 49,50/52 |
| 20 | - | 155.7 | | - | 155.6 | |
| 21 | 7.04 (1H,mult) | 116.3 | 23,25 | 7.04 (1H,mult) | 116.9 | 20,25 |
| 22 | 7.19 (1H,dd,8.1,2.6) | 129.6 | 20,24,26 | 7.01 (1H,mult) | 128.9 | 24 |
| 23 | 7.02 (1H,mult) | 120.3 | 21,,25 | 7.06 (1H,d,8.5) | 120.6 | 21,25 |
| 24 | 7.42 (1H,mult) | 127.6 | 20,22,26 | 7.46 (1H,d,8.5) | 126.8 | 22 |
| 25 | - | 131.6 | | - | 132.2 | |
| 26 | 5.44 (1H,d,8.9) | 79.4 | 11,12,22,24,25,27 | 5.53 (1H,d,7.2) | 78.9 | 11,22,24,25,27 |
| 27 | - | 155.0 | | - | 155.2 | |
| 28 | 6.82 (1H,dd,8.2,2.4) | 118.8 | 27,30,32 | 6.76 (1H,d,6.8) | ND | |
| 29 | 7.01a (1H,mult) | 127.8 | 27,31 | 6.98 (1H,mult) | ND | |
| 30 | 7.01b (1H,mult) | 113.8 | 28,32 | 6.76 (1H,d,6.8) | ND | |
| 31 | 6.93 (1H,dd,8.3,1.1) | 129.5 | 27,29,33 | 6.98 (1H,mult) | ND | |
| 32 | - | 129.6 | | - | ND | |
| 33a | 2.82 (1H,dd,14.3,11.9) | 34.9 | 3,29,32 | 2.54 (1H,mult) | 36.4 | |
| 33b | 3.18 (1H,dd,14.3,4.1) | 34.9 | 3,29,32 | 3.09 (1H,dd,13.3,3.6) | 36.4 | |
| 34a | 1.92 (1H,mult) | 20.8 | 5,35,36 | 3.75 (1H,mult) | 65.8 | 8 |
| 34b | 1.79 (1H,mult) | 20.8 | 5,6,341 | - | - | |
| 35a | 1.92 (1H,mult) | 30.3 | 5,34,36 | 0.95 (3H,d,6.2) | 19.7 | 6,34 |
| 35b | 1.79 (1H,mult) | 30.3 | 5,6,362 | - | - | |
| 36a | 3.38 (1H,mult) | 45.9 | 34 | 4.61 (1H,d,5.1) | - | 34,35 |
| 36b | 3.52 (1H,mult) | 45.9 | 34,35 | - | - | |
| 37 | 4.19 (1H,mult) | 64.8 | 8 | 1.66 (1H,mult) | 36.5 | |
| 38 | 0.85 (3H,d,6.4) | 17.4 | 9,37 | 0.75 (1H,d,6.7) | 15.1 | 9,34,39 |
| 39a | 4.93 (1H,s) | - | | 0.98 (1H,mult) | 24.4 | |
| 39b | - | - | | 1.30 (1H,mult) | 24.4 | |
| 39’ | - | - | | 0.74 (1H,d,7.3) | 15.1 | 37,39 |
| 40a | 2.31 (1H,td,13.7,7.0) | 38.3 | 14,15,41,42 | 2.33 (1H,d,13.9,7.2) | 37.9 | 14,15,41,42/44 |
| 40b | 2.49 (1H,mult) | 38.3 | 14,15,42 | 2.49 (1H,mult) | 37.9 | 41,42/44 |
| 41 | - | 126.5 | | - | 126.5 | |
| 42 | 6.78 (2H,d,8.4) | 129.5 | 40,41,42/44,43/45,46 | 6.80 (1H,d,8.4) | 129.6 | 40,42/44,46 |
| 43 | 6.51 (1H,mult) | 114.4 | 41,43/45,46 | 6.51 (1H,d,8.3) | 114.4 | 41,43/45,46 |
| 44 | 6.78 (2H,d,8.4) | 129.5 | 40,41,42/44,43/45,46 | 6.80 (1H,d,8.4) | 129.6 | 40,42/44,46 |
| 45 | 6.50 (1H,mult) | 114.4 | 41,43/45,46 | 6.51 (1H,d,8.3) | 114.4 | 41,43/45,46 |
| 46 | - | 155.5 | | - | 155.4 | |
| 47 | 9.13 (1H,s) | - | 43/45 | 9.08 (1H,s) | - | 43/45,46 |
| 48 | 6.02 (1H,d,4.5) | - | | 6.01 (1H,d,6.6) | - | 17,18,19 |
| 49 | - | 140.4 | 17,19 | - | 140.6 | |
| 50 | 7.59 (2H,mult) | 126.5 | 19,54 | 7.59 (1H,d,7.6) | 126.6 | 19,50/52,54 |
| 51 | 7.41 (1H,mult) | 127.7 | 49 | 7.40 (1H,t,7.5) | 127.9 | 49,51/53 |
| 52 | 7.59 (2H,mult) | 126.5 | 19,54 | 7.59 (1H,d,7.6) | 126.6 | 19,50/52,54 |
| 53 | 7.39 (1H,mult) | 127.7 | 49 | 7.40 (1H,t,7.5) | 127.9 | 49,51/53 |
| 54 | 7.33 (1H,t,7.2) | 127.3 | 50/52 | 7.32 (1H,t,7.3) | 127.2 | 50/52 |

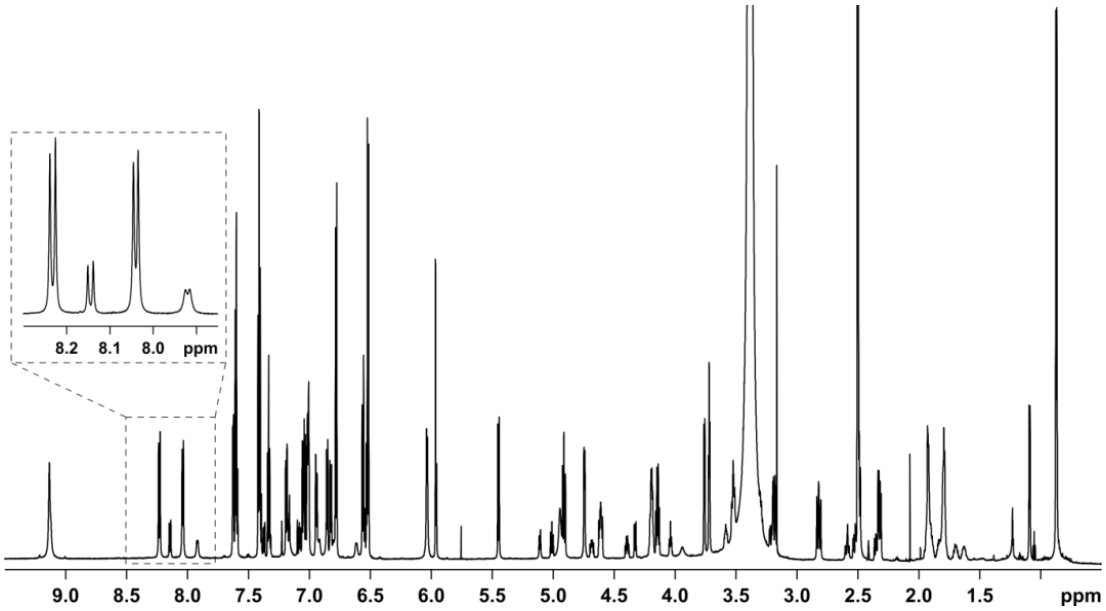



**Figure 3.5. 1D spectrum of cyclomorphosin A (CM-A) in DMSO-$d_6$ at 800 MHz. A carboxylic acid resonance at 12.76 ppm is omitted. Residual solvent and water peaks are cut off. Inset: Two slowly exchanging conformers of CM-A.**

The amino acid residues were identified as distinct spin systems from DQF-COSY and HSQC-TOCSY spectra, and these spin systems were connected by HMBC correlations and qualitative NOEs as indicated in Figure 3.6. A very low *J*-coupling constant between H-18 and H-19 made this connection difficult to assign due to a lack of correlations in e.g. the DQF-COSY experiment. The joint was thus determined from HMBC correlations, NOEs and a process of elimination of possible connectivities of the nuclei. The data is summarized in Table 3.5 and Appendix A4.

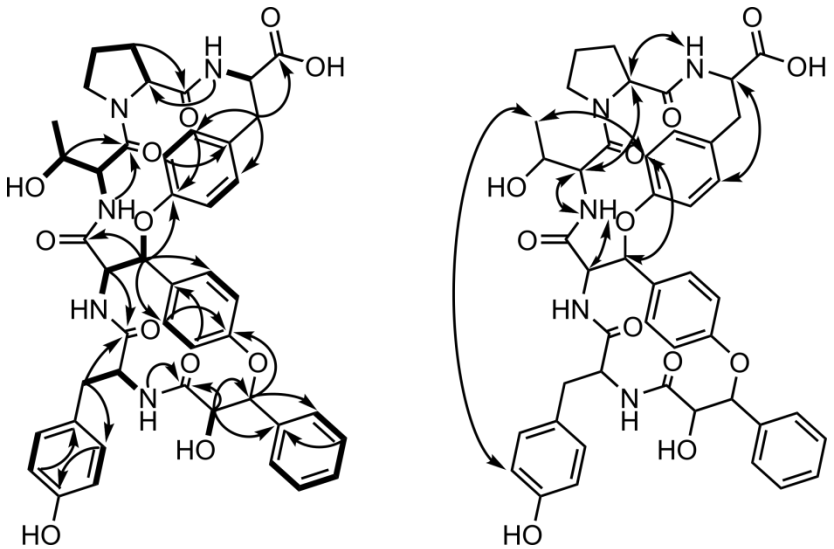


**Figure 3.6. Left: DFQ-COSY and HSQC-TOCSY (▬) spin systems and key HMBCs (→) of CM-A. Right: Key NOEs of the B ring of CM-A.**

The two NRPs are each made of five amino acids and a building block which may originate from an intermediate in the shikimate pathway.[74] Both structures exhibit a bi-cyclic motif, Figure 3.7, none of which have previously been reported from any fungal sources. Similar structures to the 14-membered ring (A), from different sets of amino acids, have previously been reported from

plants.[75–79] The 17-membered ring (B) is a completely novel motif from natural sources. Both ring structures were rather rigid due to the rigidity of aromatic systems and amide bonds, especially the 14-membered ring with a smaller ring size.

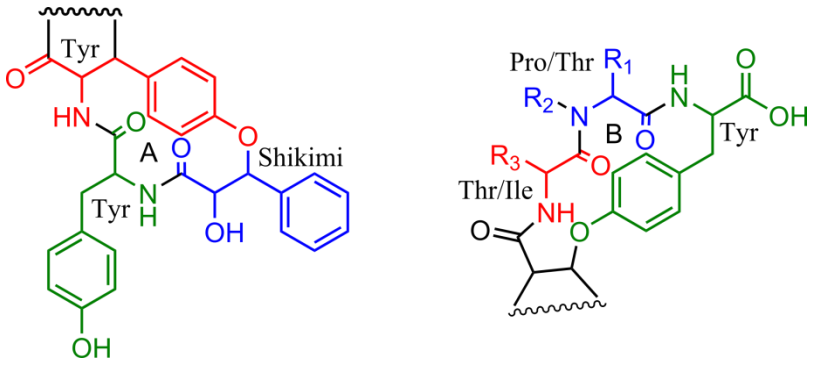



**Figure 3.7. Motif of the 14- (A) and 17-membered ring (B) of cyclomorphosin A and B. Colors illustrate the different putative building blocks.**

The structures of the cyclomorphosins have nine chiral centers in the structure. Multiple examples of D-amino acids have been reported in NRPs, including *A. homomorphus*, and the stereochemistry may not simply be ascribed to L-amino acids.[67,80] The relative orientation of the chiral centers was solved by first decimating the amount of possible diastereomers from qualitative NOE analysis, applying distances obtained from relative NOE intensities to the remaining diastereomers and identifying the best fit to the data.

Aspects of the relative stereochemistry in ring A around the oxidized tyrosine was directly read from qualitative NOEs as illustrated in Figure 3.8. NOEs indicate that H-26 and H-19 are positioned on the same face of the structure, while H-26 and H-12 are positioned on opposite faces, resulting in very different NOEs for these neighboring protons.

The differences in observed NOEs between H-12 and H-26 may not be proof of the relative stereochemistry, but considering a rather large *J*-coupling constant of 9.6 Hz between the two protons, a conformation displaying the observed NOEs while the protons are on the same face of the rings is impossible. From the HLA equation, the dihedral angles that result in a coupling constant of 9.6 Hz are ±3 and ±156 degrees. The first is not compatible with the NOE data, the latter is only possible if on opposite faces of the rings. It was also realized through computational simulations that the angle was rarely near 0° for the threo structures while it was often near ±155° for the erythro conformations.

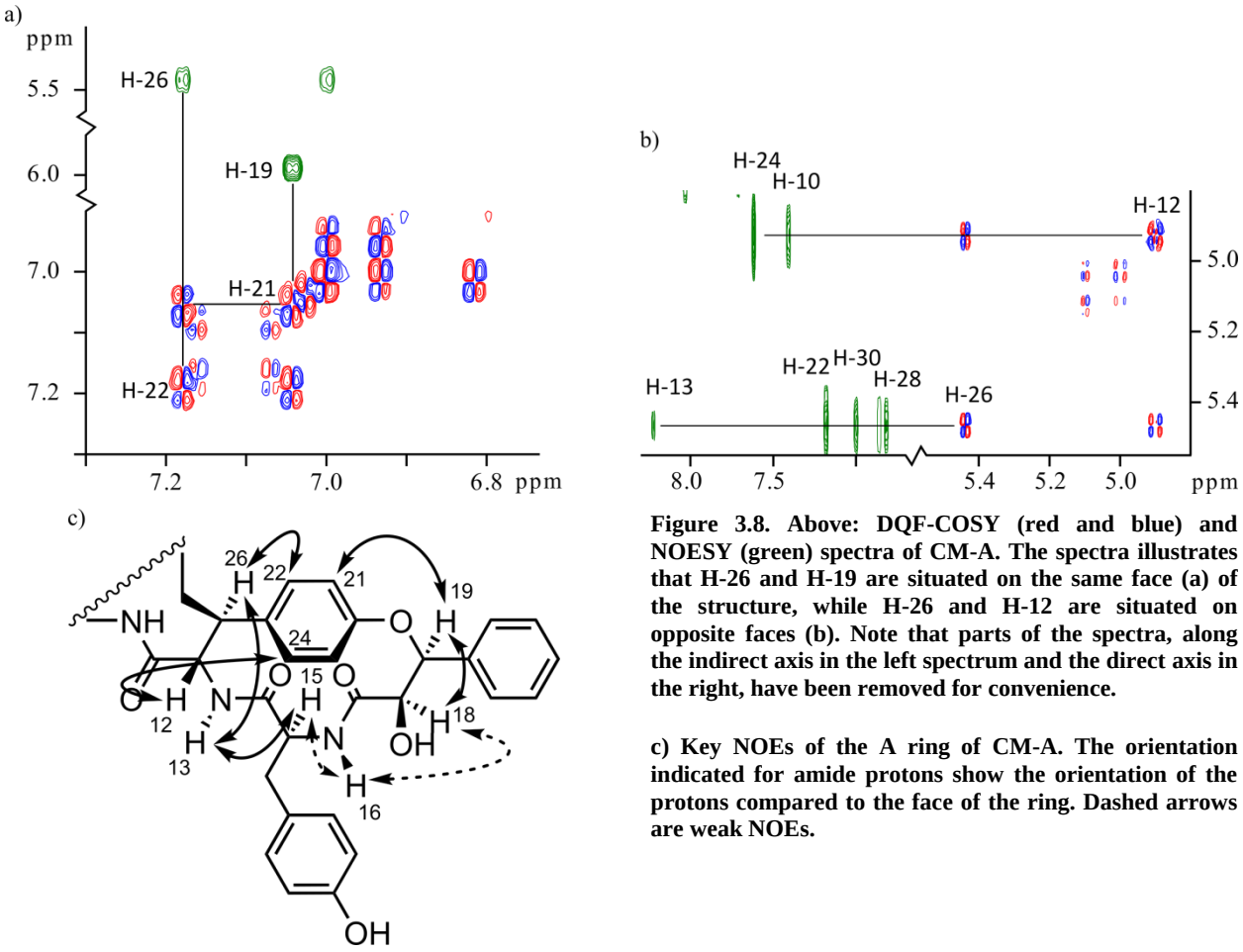



**Figure 3.8. Above: DQF-COSY (red and blue) and NOESY (green) spectra of CM-A. The spectra illustrates that H-26 and H-19 are situated on the same face (a) of the structure, while H-26 and H-12 are situated on opposite faces (b). Note that parts of the spectra, along the indirect axis in the left spectrum and the direct axis in the right, have been removed for convenience.**

**c) Key NOEs of the A ring of CM-A. The orientation indicated for amide protons show the orientation of the protons compared to the face of the ring. Dashed arrows are weak NOEs.**

In a similar fashion, it was realized that the H-19 and H-26 had to be on the same face of the ring. This led to four possible diastereomers of the A ring, while the B ring, with four unknown stereocenters, has eight possible diastereomers.

### Quantitative NOE calculations

When approaching quantitative NOE calculations of CM-A the fact that two slowly exchanging conformers were present could have been a considerable problem, as overlapping signals would be a sum of the two distinct conformational averages. The few overlapping signals present were thus excluded in the calculations.

### Ring A

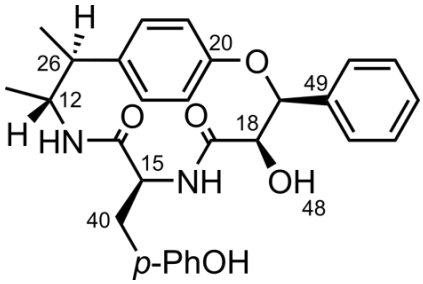
The stereochemistry of the A ring was apparent from the observations mentioned above and from simulated structures, due to differences in MAEs of the diastereomers and the rigidity for the ring. Selected distances from NOEs are presented in Table 3.6. The distances were selected as they are involved in NOEs and *J*-coupling constants which include protons in the backbone of the ring. When these are compared to 3D structures the fits are as given in Table 3.7.

**Table 3.6. Used distances and *J*-coupling constants for the determination of the stereochemistry of ring A ($\tau_m$ = 200 ms). [a]Reference distance.**

| H1 | H2 | Dist. [Å] |
|---|---|---|
| 12 | 24 | 2.57 |
| 13 | 15 | 2.18 |
| 13 | 22 | 3.10 |
| 13 | 26 | 2.37 |
| 15 | 16 | 3.13 |
| 15 | 41a | 2.69 |
| 15 | 41b | 2.80 |
| 16 | 18 | 3.26 |
| 16 | 41a | 3.04 |
| 18 | 19 | 2.51 |
| 19 | 21 | 2.02 |
| 22 | 26 | 2.30 |
| 41a | 41b | 1.75[a] |
| **H1** | **H2** | ***J* [Hz]** |
| 12 | 13 | 9.9 |
| 12 | 26 | 9.3 |
| 15 | 16 | 9.2 |
| 15 | 40a | 7.5 |
| 15 | 40b | 7.9 |
| 18 | 19 | 0.0 |

**Table 3.7. MAEs from iteratively fitting back-calculated distances to experimental data, by minimizing MAE, for ring A to MM structures from simulations of diastereomers as indicated.**

| Diastereomer | Distance [%] | | *J*-coupling [Hz] | |
|---|---|---|---|---|
| | MAE | Max | MAE | Max |
| 12*S*,15*S*,18*R*,19*S*,26*S* | 3.2 | 7.6 (H12-H24) | 0.4 | 1.1 (H15-H41a) |
| 12*S*,15*R*,18*R*,19*S*,26*S* | 6.2 | 18.5 (H16-H18) | 1.0 | 1.7 (H18-H19) |
| 12*S*,15*S*,18*S*,19*S*,26*S* | 7.8 | 33.7 (H16-H18) | 1.9 | 9.8 (H18-H19) |
| 12*S*,15*R*,18*S*,19*S*,26*S* | 6.5 | 21.4 (H19-H21) | 1.6 | 5.2 (H18-H19) |
| 12*R*,15*S*,18*R*,19*S*,26*R* | 7.6 | 23.0 (H19-H21) | 0.6 | 1.4 (H12-H13) |




It is evident that the differences in distance were not that big between the different diastereomers of CM-A. While the change of stereochemistry at position 18 was problematic due to the increase in the possible *J*-coupling constant between H-18 and H-19, the differentiation of the diastereomer of 15 was more challenging. Still the fit of the data to the *L*-diastereomer was markedly better, especially if one identifies the distances which were not observed but should yield NOEs for the *D*-diastereomer. For example the distance of H-15 to H-18, which was 2.16 Å for the 15-*D* diastereomer, from the populations obtained by fitting the data, had no corresponding observed NOE in the spectra, which correlated well with the 15-*L* diastereomer.

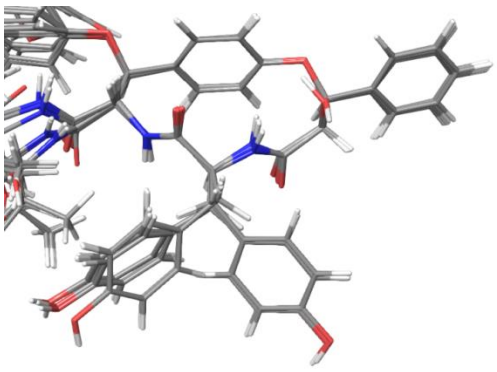


**Figure 3.9. The motif of ring A for the 12*S*,15*S*,18*R*,19*S*,26*S*-diastereomer (with the lowest MAE between experimental and theoretical data), determined from the fitting of NOE and *J*-coupling constant data to multiple 3D structures.**

Once it was established that the ring was quite rigid, another approach to determine the configuration might be to assign stereocenters at random and use constraints from NOE and *J*-coupling constant data and floating chirality. This has been proposed by Baran *et al.*[81] A similar approach could be used in the Schrödinger Suite using MacroModel. A couple of measures needed

to be taken; the chirality should be allowed to change, a mix of torsional and low-mode sampling needed to be sued, and the maximum low-mode steps needed to be quite big. In the current work, a minimum of 5 Å and a maximum of 25 Å was used (standard values are 3 Å and 6 Å, respectively). This allowed the molecule to easily switch chirality in the low mode steps, and due to the energy penalties on constraints the low energy output structures should primarily be the correct diastereomer. The distances in Table 3.6 were used as constraints with a penalty of 25 kJ/mol $Å^2$, and the *J*-coupling constants of H12-H13, H12-H26, H15-H16 and H18-H19 from Table 3.6 were translated to dihedral angles by 180 ± *n* (*n*=180±*f* (*J*)) or 0 ± *n* (*n*=0±*f* (*J*)) which is needed in MacroModel. The stereoisomer 12*S*,15*R*,18*S*,19*R*,26*S* was chosen as the starting structure, and after the simulation the major diastereomer was 12*S*,15*S*,18*R*,19*S*,26*S* which constituted 67 % of the structures determined in a window of 50 kJ/mol and 95 % in a window of 25 kJ/mol from the determined energy minimum (3,111 total structures), see Table 3.8 for more details.

**Table 3.8. Result of floating chirality simulation, as described in the text, of ring A using distance and dihedral constraints from Table 3.6, using different energy cut-offs.**

| Diastereomer | Population [%] | | |
|---|---|---|---|
| | 50 kJ/mol | 25 kJ/mol | 10 kJ/mol |
| 12*S*,15*S*,18*R*,19*S*,26*S* | 67.0 | 95.0 | 100.0 |
| 12*S*,15*R*,18*R*,19*S*,26*S* | 32.7 | 5.0 | 0.0 |
| 12*S*,15*S*,18*S*,19*S*,26*S* | 0.0 | 0.0 | 0.0 |
| 12*S*,15*R*,18*S*,19*S*,26*S* | 0.3 | 0.0 | 0.0 |
| 12*R*,15*S*,18*R*,19*S*,26*R* | 0.0 | 0.0 | 0.0 |

**Ring B**

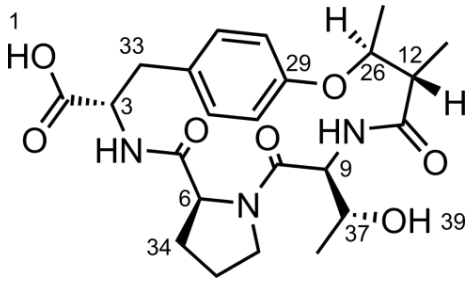
The B ring was more difficult to solve due to a higher degree of flexibility originating from the reduced rigidity, and a larger conformational space for the larger ring size. Since the diastereomers of ring A were solved, the three remaining chiral centers in the ring were varied and the best minimum MAE identified. The assignments of the diastereotopic protons of C-33 in the ring were allowed to interchange to best fit the data. Again, only select distances were used following the same selection criteria as for ring A. Also, distances between the protons of proline were ignored as these are shared between all possible conformers due to ring strain in the five-membered ring, and major resonance overlaps between protons in the ring.

**Table 3.9. Used distances and *J*-coupling constants for the determination of the stereochemistry of ring B ($\tau_m$ = 200 ms). [a]Reference distance.**

| H1 | H2 | Dist. [Å] |
|---|---|---|
| 3 | 29 | 2.42 |
| 3 | 31 | 2.94 |
| 3 | 33b | 2.42 |
| 4 | 6 | 2.34 |
| 4 | 29/31 | 3.16 |
| 4 | 33a | 2.84 |
| 4 | 33b | 3.76 |
| 6 | 9 | 2.06 |
| 6 | 37 | 2.73 |
| 9 | 10 | 2.71 |
| 9 | 37 | 2.53 |
| 10 | 12 | 2.22 |
| 10 | 13 | 2.96 |
| 10 | 38 | 2.98 |
| 26 | 30 | 2.07 |
| 29 | 33a | 2.64 |
| 29 | 33b | 2.55 |
| 31 | 33a | 2.27 |
| 33a | 33b[a] | 1.75 |
| 37 | 38 | 2.56 |
| H1 | H2 | *J* [Hz] |
| 3 | 4 | 8.8 |
| 3 | 33a | 11.7 |
| 9 | 10 | 4.5 |
| 9 | 37 | 4.4 |

**Table 3.10. MAEs from iteratively fitting back-calculated distances to experimental data, by minimizing MAE, for ring B to MM structures from simulations of diastereomers as indicated.**

| Diastereomer | Distance [%] | | *J*-coupling [Hz] | |
|---|---|---|---|---|
| | MAE | Max | MAE | Max |
| 3*S*,6*S*,9*S*,12*S*,26*S* | 4.6 | 11.4 (H29-H33b) | 0.0 | 0.1 (H9-H10) |
| 3*R*,6*S*,9*S*,12*S*,26*S* | 6.6 | 21.1 (H29-H33b) | 1.2 | 1.2 (H3-H4) |
| 3*S*,6*R*,9*S*,12*S*,26*S* | 6.8 | 15.9 (H4-H6) | 0.0 | 0.5 (H3-H33a) |
| 3*S*,6*S*,9*R*,12*S*,26*S* | 6.4 | 21.2 (H6-H9) | 0.2 | 0.2 (H3-H4) |
| 3*R*,6*R*,9*S*,12*S*,26*S* | 8.2 | 25.2 (H3-H29) | 0.5 | 0.5 (H3-H4) |
| 3*R*,6*S*,9*R*,12*S*,26*S* | 6.6 | 20 (H26-H30) | 0.3 | 0.3 (H3-H4) |
| 3*S*,6*R*,9*R*,12*S*,26*S* | 6.0 | 17.9 (H6-H9) | 0.8 | 0.8 (H3-H4) |
| 3*R*,6*R*,9*R*,12*S*,26*S* | 6.7 | 17.9 (H26-H30) | 0.0 | 0 (-) |




The resulting MAEs from all diastereomers were almost equal, and it was found more illustrious to compare the increase or decrease in correlation to the individual distances of the data. This is done in Table 3.11 as the difference between back-calculated and experimental distances (in %) and *J*-coupling constants (in Hz) for the relevant diastereomers. The all *S* diastereomer had the lowest MAE, and interchanging chiral centers led to several rather big errors in many distances and *J*-coupling constants between experiments and calculations. The two diastereomers 3*S*,6*R*,9*R*,12*S*,26*S* and 3*R*,6*R*,9*R*,12*S*,26*S* were the hardest diastereomers to dismiss by the data. The first one (3*S*,6*R*,9*R*,12*S*,26*S* ) could be dismissed from the increase in the distance between H6 and H9 which translates to two alpha protons being too far apart in space. The latter diastereomer (3*R*,6*R*,9*R*,12*S*,26*S*) had an increase in the error of H26 distance to the aromatic ring in addition to an error in the H6-H9 distance.

**Table 3.11. Error in distances/*J*-coupling constants in %/Hz compared to the 3*S*,6*S*,9*S*,12*S*,26*S*-diastereomer. Diastereomers varied as indicated. Red/blue indicates an increase/decrease in error of more than 5 %/0.5 Hz.**

| Change in distance errors [%] | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| H1 | H2 | 3*R* | 6*R* | 9*R* | 3*R*6*R* | 3*R*9*R* | 6*R*9*R* | 3*R*6*R*9*R* |
| 3 | 29 | 10.2 | -1.9 | 8.5 | 18.6 | 6.0 | -0.8 | 5.0 |
| 3 | 31 | 2.3 | -0.4 | 3.0 | 9.7 | -0.5 | -0.1 | 1.7 |
| 3 | 33b | -1.7 | 1.1 | -2.3 | -2.5 | -1.0 | -1.9 | -0.2 |
| 4 | 6 | 6.7 | 14.4 | 2.9 | 0.6 | 11.0 | 4.8 | 1.4 |
| 4 | 33a | 5.1 | 4.2 | 4.3 | 4.8 | 2.4 | 1.8 | 3.4 |
| 4 | 33b | -1.9 | -1.8 | -1.2 | -0.4 | -0.9 | -0.8 | -1.2 |
| 6 | 9 | -0.9 | 6.3 | 11.7 | 13.3 | 2.7 | 8.4 | 6.4 |
| 9 | 10 | -1.2 | -1.2 | 0.5 | -2.5 | -1.3 | 0.0 | 0.1 |
| 9 | 37 | 1.7 | 0.1 | 1.7 | 0.6 | 3.4 | 5.7 | 4.5 |
| 10 | 12 | -1.0 | 2.0 | -0.3 | 4.7 | 1.5 | 3.5 | 3.2 |
| 10 | 13 | 1.4 | 15.1 | 3.8 | 0.9 | 2.4 | -0.3 | 2.1 |
| 10 | 38 | -2.5 | 1.7 | 2.2 | 4.4 | -4.5 | -2.4 | 0.3 |
| 26 | 30 | 5.9 | 1.2 | -0.9 | 12.5 | 13.7 | 4.4 | 11.5 |
| 29 | 33a | -6.3 | 1.5 | 0.6 | -4.9 | -3.8 | 1.6 | -5.3 |
| 29 | 33b | 9.7 | -2.0 | -3.2 | -0.8 | 0.4 | 0.0 | 1.2 |
| 31 | 33a | 9.1 | 0.5 | 0.7 | 5.7 | 5.0 | 1.2 | 3.2 |
| 37 | 38 | 0.0 | -0.1 | 0.2 | 0.1 | 0.2 | -0.2 | 0.1 |
| **Change in *J*-coupling constants [Hz]** | | | | | | | | |
| H1 | H2 | 3*R* | 6*R* | 9*R* | 3*R*6*R* | 3*R*9*R* | 6*R*9*R* | 3*R*6*R*9*R* |
| 3 | 4 | 1.2 | 0.0 | 0.2 | 0.5 | 0.3 | 0.8 | 0.0 |
| 3 | 33a | 0.0 | 0.5 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 |
| 9 | 10 | 0.0 | -0.1 | -0.1 | -0.1 | -0.1 | -0.1 | -0.1 |
| 9 | 37 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

A possible path to gain more certainty of the stereochemistry was assume a better accuracy in the relative energies of the MMFFs force field, and limit the amount of structures based on energies. If the window was shrunk, from including structures within 50 kJ/mol of the identified minimum for each conformation, to 15 kJ/mol (which translates to a Boltzmann distribution population of ~0.2 %), the stereochemistry is much more easily solved, as seen in Table 3.12.

**Table 3.12. Fit, obtained by iteratively minimizing MAE, of NOE derived distances for ring B to those of MM structures from simulations of diastereomers. Structures within 30 kJ/mol of the global minimum used.**

| Diastereomer | Distance [%] | | *J*-coupling [Hz] | |
|---|---|---|---|---|
| | MAE | Max | MAE | Max |
| 3*S*,6*S*,9*S*,12*S*,26*S* | 5.0 | 14.2 (H10-H38) | 0.1 | 0.3 (H9-H10) |
| 3*R*,6*S*,9*S*,12*S*,26*S* | 11.4 | 38.4 (H10-H13) | 1.1 | 3.3 (H9-H10) |
| 3*S*,6*R*,9*S*,12*S*,26*S* | 16.3 | 120.2 (H6-H9) | 1.0 | 3.6 (H9-H10) |
| 3*S*,6*S*,9*R*,12*S*,26*S* | 16.7 | 122.1 (H6-H9) | 0.9 | 3.4 (H9-H10) |
| 3*R*,6*R*,9*S*,12*S*,26*S* | 16.0 | 121.6 (H6-H9) | 0.9 | 3.6 (H9-H10) |
| 3*R*,6*S*,9*R*,12*S*,26*S* | 14.5 | 120.4 (H6-H9) | 0.9 | 2.6 (H9-H10) |
| 3*S*,6*R*,9*R*,12*S*,26*S* | 8.7 | 20.8 (H6-H9) | 1.2 | 2.5 (H9-H10) |
| 3*R*,6*R*,9*R*,12*S*,26*S* | 8.5 | 30.2 (H26-H30) | 0.2 | 0.7 (H3-H4) |

The resulting structure consisted of all *L*-amino acids. To verify the assignment of *L*-amino acids, Marfey's analysis was performed and the proline was determine as *L*-proline.[82,83] Proline was chosen, as acidic amino acids are less suitable for Marfey's analysis due to poor separation.[82]

## Full structure

The fitting of 3D structures to the full set of NMR data for the proposed diastereomer is illustrated in Table 3.13, Table 3.14 and Figure 3.10, and a representation of the structural space occupied (by multiple structures) is found in Appendix A4. The MAE was comparable to the MAE obtained from the much more rigid homomorphosin A, which was slightly surprising due to the more flexible nature of the compound. This could be a sign of overfitting, which is a problem when fitting multiple structures.

**Table 3.13. Comparison of the distances from NOE intensities and 3D structures using multiple structures from MM ($\tau_m$ = 200 ms). All distances are in Å. [a]Reference distance.**

| H1 | H2 | Exp. dist. | 3D Dist. | Diff (Å) | Diff (%) |
|---|---|---|---|---|---|
| 3 | 29 | 2.52 | 2.52 | 0.00 | 0.2 |
| 3 | 31 | 3.06 | 3.07 | 0.01 | 0.4 |
| 3 | 33b | 2.52 | 2.51 | 0.01 | 0.5 |
| 4 | 6 | 2.43 | 2.34 | 0.09 | 3.8 |
| 4 | 33a | 2.96 | 2.83 | 0.14 | 4.6 |
| 4 | 33b | 3.91 | 3.85 | 0.06 | 1.5 |
| 4 | 35b | 3.29 | 3.08 | 0.22 | 6.5 |
| 4 | 36b | 3.46 | 3.47 | 0.02 | 0.5 |
| 6 | 9 | 2.15 | 2.31 | 0.16 | 7.3 |
| 6 | 34a | 2.39 | 2.35 | 0.04 | 1.7 |
| 6 | 34b | 2.59 | 2.73 | 0.14 | 5.5 |
| 9 | 10 | 2.82 | 2.80 | 0.01 | 0.5 |
| 9 | 37 | 2.64 | 2.69 | 0.05 | 1.8 |
| 10 | 12 | 2.31 | 2.17 | 0.14 | 6.1 |
| 10 | 13 | 3.08 | 3.01 | 0.07 | 2.2 |
| 10 | 38 | 3.11 | 3.32 | 0.21 | 6.7 |
| 12 | 24 | 2.57 | 2.75 | 0.19 | 7.3 |
| 13 | 15 | 2.18 | 2.13 | 0.05 | 2.5 |
| 13 | 22 | 3.10 | 3.23 | 0.13 | 4.2 |
| 13 | 26 | 2.75 | 2.92 | 0.17 | 6.0 |
| 15 | 16 | 3.13 | 2.99 | 0.15 | 4.6 |
| 15 | 40a | 2.66 | 2.73 | 0.06 | 2.3 |
| 15 | 40b | 2.80 | 2.76 | 0.05 | 1.6 |
| 16 | 18 | 3.26 | 3.29 | 0.02 | 0.6 |
| 16 | 40a | 3.04 | 2.86 | 0.18 | 5.8 |
| 18 | 19 | 2.51 | 2.64 | 0.50 | 5.2 |
| 19 | 21 | 2.02 | 2.14 | 0.13 | 5.9 |
| 22 | 26 | 2.30 | 2.32 | 0.12 | 0.9 |
| 26 | 30 | 2.16 | 2.19 | 0.02 | 1.2 |
| 29 | 33a | 2.75 | 2.92 | 0.03 | 6.3 |
| 29 | 33b | 2.65 | 2.80 | 0.17 | 5.6 |
| 31 | 33a | 2.36 | 2.48 | 0.15 | 5.0 |
| 33a | 33b | 1.82 | 1.75 | 0.12 | 4.0 |
| 34b | 36b | 2.40 | 2.43 | 0.07 | 1.3 |
| 35b | 36a | 2.30 | 2.42 | 0.03 | 4.9 |
| 37 | 38 | 2.67 | 2.62 | 0.11 | 1.7 |
| 38 | 43/45 | 3.60 | 3.64 | 0.05 | 1.1 |
| 40a | 40b | 1.75[a] | 1.75 | 0.05 | 0.0 |
| | | | MAE | 0.10 | 3.4 |

**Table 3.14. Comparison of experimental *J*-coupling constants and calculated constants by the Karplus or HLA equation in Hz, from 3D structures using multiple structures from MM.**

| H1 | H2 | *J* exp. | *J* calc. | Diff |
|---|---|---|---|---|
| 3 | 4 | 8.8 | 8.6 | 0.2 |
| 3 | 33a | 11.7 | 11.8 | 0.1 |
| 9 | 10 | 4.5 | 4.8 | 0.3 |
| 9 | 37 | 4.4 | 4.4 | 0.0 |
| 12 | 13 | 9.9 | 9.4 | 0.5 |
| 12 | 26 | 9.3 | 9.7 | 0.4 |
| 15 | 16 | 9.2 | 9.1 | 0.0 |
| 15 | 40a | 7.5 | 7.4 | 0.0 |
| 15 | 40b | 7.9 | 7.9 | 0.0 |
| 18 | 19 | 0 | 0.8 | 0.8 |

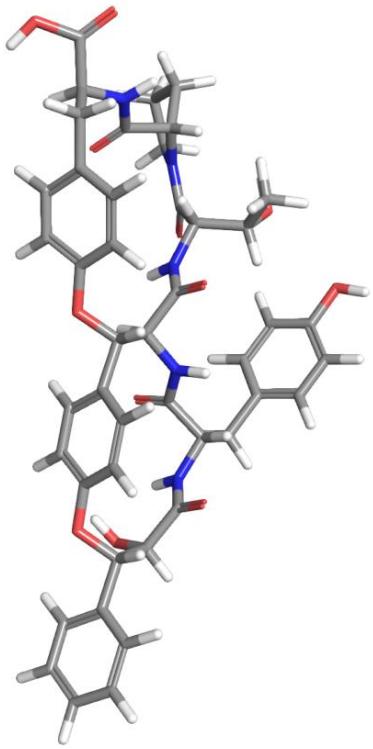


**Figure 3.10. Highest populated structure from the NOE and *J*-coupling constant data. Used as a representative structure for the structural space inhabited by cyclomorphosin A, suggested from NOE and *J*-coupling constant data.**

The stereochemistry of the novel NRP was thus solved based on NOEs and *J*-coupling constants. While the A ring is considered solved beyond doubt, the B ring proved more troublesome. The obtained solution yields by far the best correlation to the experimental data though, leaving it as the best possible solution. The problem would not have been properly addressed using a single structure due to the degree of flexibility and resulting high differences of data to 3D structures for all diastereomers leading to poor discrimination of their consistency with the NMR data.

Due to a very limited amount of compound, a similar analysis of CM-B was not possible, and not even all $^{13}C$ resonances were observed in the spectra. The assignment is thus based on the very similar chemical shifts, spin systems and NOEs at long mixing times. The A ring is deemed of identical, or fully reversed, stereochemistry due to very similar *J*-coupling constants and chemical shifts, while the stereochemistry of the B ring was not solved due to differences in the amino acids and an increased flexibility from substituting a proline to a threonine. The minor conformer observed in the CM-B spectra was in the ratio 10:1, which is contributed to relieved strain in the B ring from removal of the proline, and it is suggested that the minor conformation is based in changes in the B ring for both structures.

### 3.1.6 Homomorphosterol

The novel sterol homomorphosterol, see Figure 3.11, was isolated from *A. homomorphus*. In the work of the Master Thesis "Chemistry of Black Aspergilli" the stereochemistry was wrongfully assigned, and the determination of the right diastereomer is thus included here. This is a good case of usage of quantum chemical calculations in solving the stereochemistry of rigid compounds.

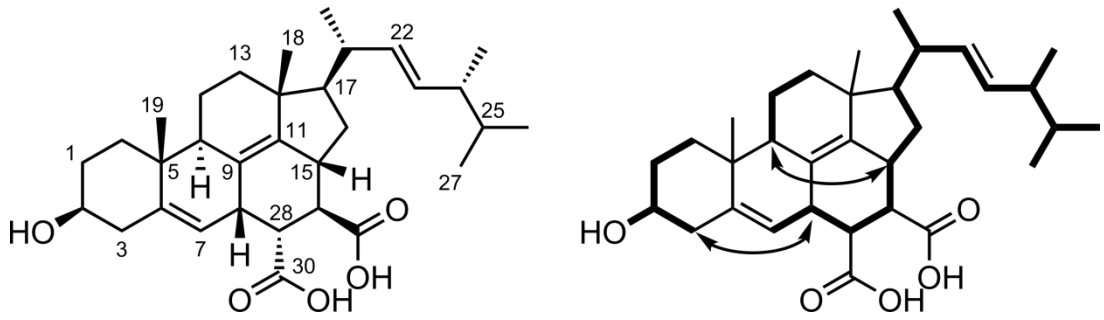



**Figure 3.11. The structure of homomorphosterol (left) and the spin systems as determined by DQF-COSY and HSQC-TOCSY (right). Arrows indicate observed $^{5}J$-coupling constants.**

As indicated in Figure 3.11 large $^{5}J$-coupling constants of 6 and 5 Hz were observed which has also previously been reported for various sterols.[84] This feature should be possible to address with the usage of computational chemistry. Featured important HMBC correlations and NOEs are depicted in Figure 3.12, and while the stereochemistry is in principle solvable from qualitative NOE analysis more certainty was desired. It should be noted that only data with a mixing time of 800 ms were available, and with no build-up curve and probably a too long mixing time, quantitative NOE analysis was not pursued due to a probable, ultimately unknown, degree of spin diffusion.

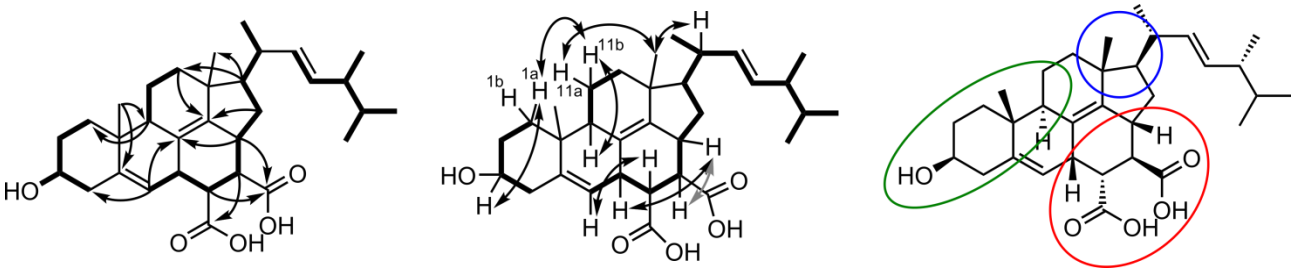


**Figure 3.12. Selected HMBC connectivities (left) and NOEs (center) of homomorphosterol. The grey arrow indicates that scalar coupling induced artifacts in the NOESY spectra made it impossible to judge if an NOE was present. Right: The three stereo-clusters in the cyclic part of homomorphosterol. Stereocenters within the green and blue circles are thought equal to the stereochemistry of ergosterol.**

An approach to solve the stereochemistry of rigid compounds is by DFT calculation of chemical shifts and *J*-coupling constants. The utilization of chemical shifts and a probability function was deemed useful in this case.[58] It should be noted that homomorphosterol features three clusters of stereocenters as indicated in Figure 3.12. It was proposed that the biosynthesis of homomorphosterol involved a Diels-Alder like reaction from an oxidized ergosterol and two of the clusters, blue and green in Figure 3.12, are structurally most likely identical to the structure of ergosterol.[85] A similar compound was previously reported where a similar biosynthesis as the one suggested in Figure 3.13 was proposed, utilizing maleimide instead of maleic or fumaric acid.[86]

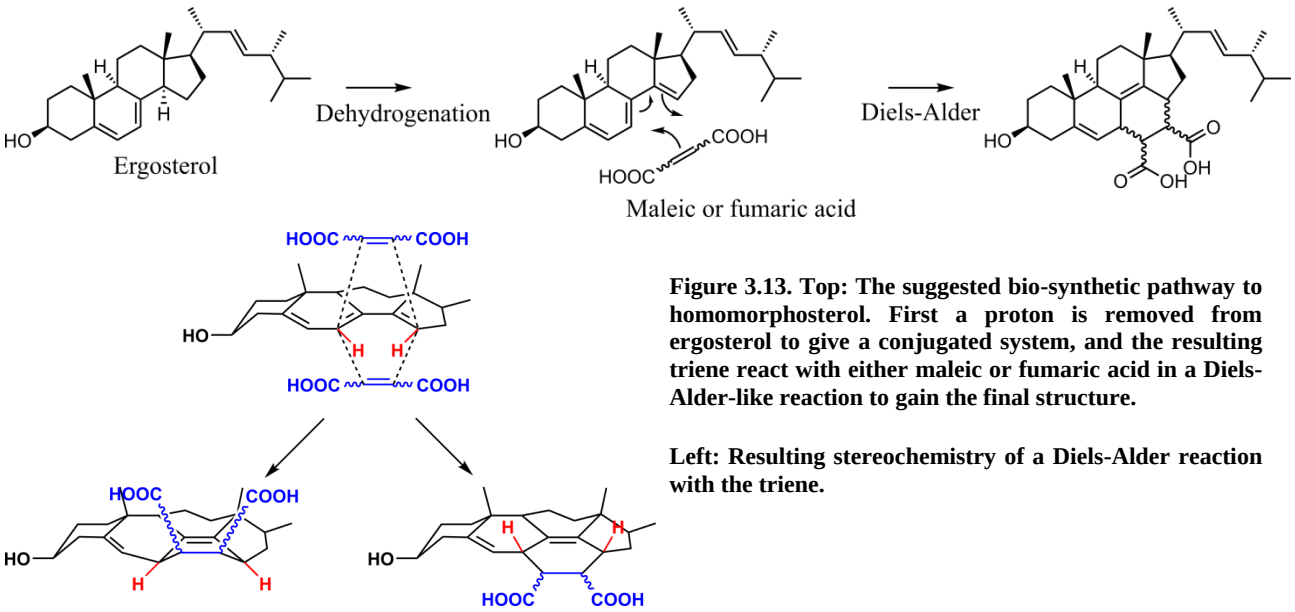



**Figure 3.13. Top: The suggested bio-synthetic pathway to homomorphosterol. First a proton is removed from ergosterol to give a conjugated system, and the resulting triene react with either maleic or fumaric acid in a Diels-Alder-like reaction to gain the final structure.**

**Left: Resulting stereochemistry of a Diels-Alder reaction with the triene.**

This leaves essentially three stereocenters to be solved; H-3 and H-6 (which are coupled), H-30 and H-31. DFT-optimized structures of truncated diastereomers, with the flexible side-chain at C-17 changed to a methyl group as given in Figure 3.14, were used in a DP4 probability analysis, as calculated by equation (3.1).[58]

$$P(i|\delta_N) = \frac{\prod_{k=1}^{N} 1 - T^{v}\left(\left|\delta_{scaled,k}^{-i} - \delta_{exp,k} - \mu\right|/\sigma\right)}{\sum_{j=1}^{m}\prod_{k=1}^{N} 1 - T^{v}\left(\left|\delta_{scaled,k}^{-i} - \delta_{exp,k} - \mu\right|/\sigma\right)} \qquad (3.1)$$

Where $\mu$ = 0, $v$ = 11.38 ($^{13}$C) or 14.18 ($^{1}$H) and $\sigma$ = 2.306 ($^{13}$C) or 0.185 ($^{1}$H) ppm.

Only a single conformer was identified for each diastereomer due to the highly rigid nature of the compound. The rigidity also results in rather big differences in the spatial structure for each diastereomer as illustrated in Figure 3.14, which accordingly yields differences in theoretical chemical shifts and *J*-coupling constants.

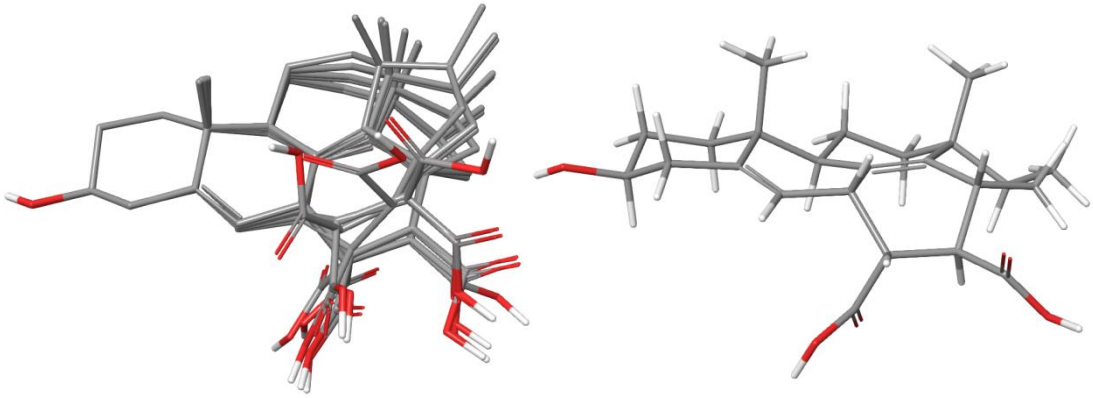


**Figure 3.14. Left: Overlay of the eight stereoisomers. A quite big differences in the occupied structural space is observed. Right: The stereoisomer (*RRRS*) which showed the best correlation to the available data. All structures are truncated at C-18 to focus on the rigid core.**

The chemical shifts are collected in Appendix A4. The comparison of the DP4 probability of experimental versus theoretical chemical shifts of the diastereomers is given in Table 3.15, which clearly shows that the (RRRS) structure was favored. Especially the proton chemical shifts were discriminative which was surprising since the carbon chemical shifts hold a larger chemical shift range and thus could potentially be more discerning. This may be due to the added electronic complexity when modelling the $^{13}$C nuclei compared to the much simpler protons. The MAE for the (RRRS) diastereomer was 0.12 ppm for $^{1}$H and 1.68 ppm for $^{13}$C chemical shifts which resembles the 20 times larger chemical shift range of $^{13}$C and might explain why the proton chemical shifts were more discerning. This is in good agreement with findings from the literature.[47,54,58]

**Table 3.15. DP4 probability analysis of stereoisomers of homomorphosterol (values in %). The first row covers $^{13}$C and $^{1}$H chemical shifts. The best and worst fit to the data is marked in blue and red respectively. Shielding tensors calculated by MPW1PW91/6-311+G(2d,p) using GIAO from B3LYP/6-31g(d) optimized structures.**

| | (*RRSS*) | (*RRRS*) | (*RSRS*) | (*RSSS*) | (*SRSR*) | (*SRRR*) | (*SSSR*) | (*SSRR*) |
|---|---|---|---|---|---|---|---|---|
| DP4 $^{13}$C/$^{1}$H | 0 | 99.9 | 0.1 | 0 | 0 | 0 | 0 | 0 |
| DP4 $^{13}$C | 0.1 | 58.8 | 11.5 | 8.1 | 0.7 | 3.9 | 16.8 | 0 |
| DP4 $^{1}$H | 0 | 99.7 | 0.3 | 0 | 0 | 0 | 0 | 0 |

Also the *J*-coupling constants, including the large long-range constants, were investigated and gave the same conclusion for the stereochemistry, see Table 3.16. As the (RRRS) structure did not violate any qualitative NOE correlations, this structure was assumed to be valid. For more discussion on the DFT calculation of *J*-coupling constants see Section 4.2.1.

**Table 3.16. Experimental and theoretical *J*-coupling constants in Hz for stereoisomers of homomorphosterol. Indented numbers are the absolute difference compared to the measured couplings. Coupling constants were measured from a 1D $^1$H spectrum or DQF-COSY(*). The best and worst fits to the data are marked in blue and red, respectively. Theoretical coupling constants calculated by B3LYP/6-31g(d,p) u+1s from B3LYP/6-31g(d) optimized structures.**

| *J* [Hz] | Meas. | (*RRSS*) | (*RRRS*) | (*RSRS*) | (*RSSS*) | (*SRSR*) | (*SRRR*) | (*SSSR*) | (*SSRR*) |
|---|---|---|---|---|---|---|---|---|---|
| H3-H31 | 9.9 | 9.7 | 10.0 | 10.9 | 9.8 | 11.5 | 9.2 | 11.3 | 4.2 |
| | | *0.4* | *0.1* | *0.8* | *0.3* | *1.4* | *0.9* | *1.2* | *5.9* |
| H6-H30 | 6.9 | 4.2 | 7.7 | 11.0 | 10.3 | 10.8 | 10.8 | 8.5 | 10.8 |
| | | *2.7* | *0.8* | *4.1* | *3.4* | *3.9* | *3.9* | *1.6* | *3.9* |
| H30-H31 | 5.7 | 4.4 | 5.3 | 11.4 | 12.0 | 11.5 | 7.4 | 6.2 | 4.3 |
| | | *1.3* | *0.4* | *5.7* | *6.3* | *5.8* | *1.7* | *0.5* | *1.4* |
| H6-H9a* | 6.0 | 6.9 | 6.1 | 6.1 | 7.0 | 3.5 | 3.5 | 4.0 | 4.9 |
| | | *0.9* | *0.1* | *0.1* | *1.0* | *2.5* | *2.5* | *2.0* | *1.1* |
| H3-H14* | 5.0 | 5.2 | 4.6 | 4.8 | 5.1 | 5.2 | 5.4 | 5.3 | 4.3 |
| | | *0.2* | *0.4* | *0.2* | *0.1* | *0.2* | *0.4* | *0.3* | *0.7* |

In conclusion the stereochemistry of a novel steroid was determined from calculations of NMR properties, based on *a priori* knowledge of the dominating biosynthetic pathway of steroids in fungi, and qualitative NOEs. Note that the a priori knowledge can lead to the absolute stereochemistry if the assumptions are correct. If incorrect, the relative stereochemistry is still upheld from qualitative NOEs and the calculations above. The rigid structure was here a necessity to avoid too many structures per stereoisomer, though averaging is possible.[54] This is a prime example of how DFT computation of NMR observables can be a crucial method in solving unknown structures, a method which will only find increased usage with the introduction of more observables, e.g. in the form of more long-range coupling constants as described in Chapter 4.

## 3.2 Other natural products

In addition to the compounds isolated from *A. homomorphus*, other natural compounds were investigated. Some of these are included briefly in the following.

### 3.2.1 Aculenes

The aculenes, as well as the epi-10,23-dihydro-24,25-dehydroaflavine discussed in the next section, are small molecules isolated by Dr. Lene M. Pedersen, for which the 3D structure were determined from *J*-coupling constants, structural calculations and select qualitative NOEs.[87] The compounds were isolated from another black aspergillus; the fungus *A. aculeatus*.[69,70]

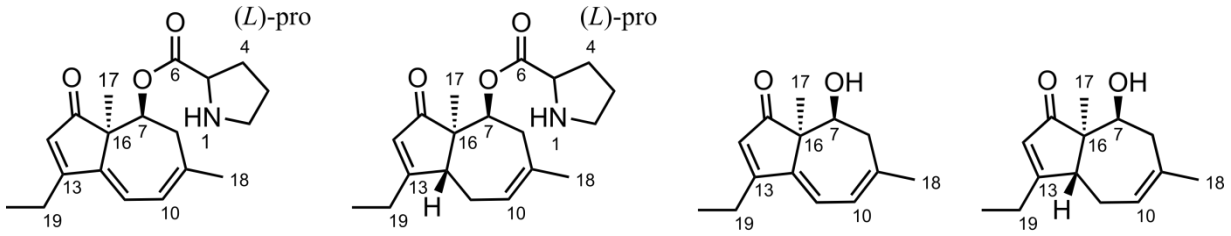



**Figure 3.15. The structures of aculene A-D with the determined stereochemistry shown.**

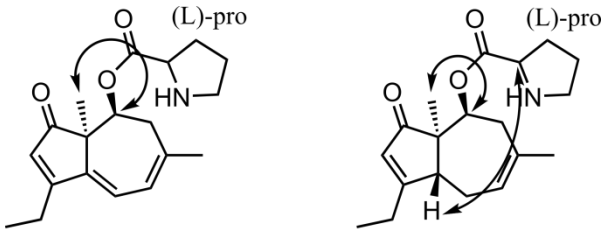
The stereochemistry of the aculenes was determined mostly from *J*-coupling constants. It was evident that the two stereocenters in the 7-membered ring of aculene A needed to be (*R*)/(*S*) from theoretical *J*-coupling constants, see Table 3.17. This is independent of the assignment of the

diastereotopic protons at C-8, as an exchange of conformation would not improve the fit to the data of any stereoisomer. From the *J*-coupling constants it was suggested that the minor conformer was lowly populated. The stereochemistry of the proline was solved by Marfey's reagent, but as it may rotate freely around the C7-O bond, it was not possible to correlate the stereocenter to the remaining structure.

**Table 3.17. Measured and theoretical *J*-coupling constants in Hz for the possible stereoisomers of the aculene A. Theoretical coupling constants calculated by B3LYP/6-31g(d,p) u+1s from B3LYP/6-31g(d) optimized structures.**

| H1 | H2 | Exp. | (RS) or (SR) | | (RR) or (SS) | |
|---|---|---|---|---|---|---|
| | | | major | minor | major | minor |
| 7 | 8 | 4.4 | 4.5 | 11.0 | 9.3 | 11.3 |
| 7 | 8' | 2.8 | 2.6 | 5.2 | 0.6 | 5.9 |

Due to a quite plausible biosynthetic relationship of the aculene A, B and C, and similar *J*-coupling constants and NOEs, the stereochemistry was thought to be identical. The remaining stereocenter of aculene B was determined qualitatively from NOEs. It should be noted that aculene D was not isolated but suggested from HRMS data, and thus no NMR data were available in support.




Aculene A Aculene B

**Figure 3.16. Qualitative NOEs used for solving the structures of aculene A and B ($\tau_m$ = 800 ms).**

## 3.2.2 Epi-10,23-dihydro-24,25-dehydroaflavine

The stereochemistry of epi-10,23-dihydro-24,25-dehydroaflavinine was solved by the use of NOEs and *J*-couplings. The relative stereochemistry of epi-10,23-dihydro-24,25-dehydroaflavinine was established by performing conformational searches on different diastereomers and by comparing observed distances to the back-calculated distances from ISPA, and by comparison of experimental $^{3}J$-couplings to HLA back-calculations. The relative stereochemistry was determined to be equal to that of the original 10,23-dihydro-24,25-dehydroaflavinine,[88,89] but as the optical rotation was of the opposite sign, compared to the literature, the enantiomer was most likely isolated. The structure that resulted in the lowest MAE was further optimized by HF and DFT to give the reported structure. Again a lower error was obtained if the rotating groups, the isopropylene and the alcohol, were allowed to rotate, see Table 3.19 and Table 3.20, as the inclusion of multiple structures eliminated most of the observed errors when using only to a single structure. It should be noted that while the single structure was optimized to a DFT level, the multiple structures are used directly from the force field simulations, accounting for some of the differences in the fit of the rigid parts and in the reference distance.

By using the ISPA approach it was possible to unambiguously assign the diastereomer of the compound. This might have been possible using qualitative data and a model, which was how the enantiomer was previously solved,[89] but the complicated geometry of the structure made ISPA an easier approach, where one did not have to guess whether the stereochemistry was actually the best possibility. While the alcohol and vinyl groups were rotating in the identified conformer populations, the orientation of the indole ring was fixed, in good correlation to the force field energies, where a reorientation of the indole ring of 180° gave rise to an approximate rise in energy of 16 kJ/mol (a population of ~0.16 %). Even with the discriminative properties of NOEs this population should not be detected. While the force field energies may be prone to errors, the obtained population of fitting the rotamers correlated pretty well to that of the energies, as seen in Table 3.18.

**Table 3.18. Population of the minimum energy conformation (1) and different rotamers of C-23 (2), C-19 (3) and both (4). Populations from force field energies and NOE data.**

| Conformer | Rel. E [kJ/mol] | Rel. FF pop. [%] | Rel. NOE pop. [%] |
|---|---|---|---|
| 1 | 0 | 46 | 56 |
| 2 | 1.26 | 28 | 12 |
| 3 | 2.65 | 16 | 23 |
| 4 | 3.89 | 10 | 9 |

In order to confirm the assignment of the diastereomer a constrained conformer search with floating chirality, using only data between methine and methyl protons was also conducted, following the procedure given for cyclomorphosin A, which resulted in only the given diastereomer in a 50 kJ/mol window.

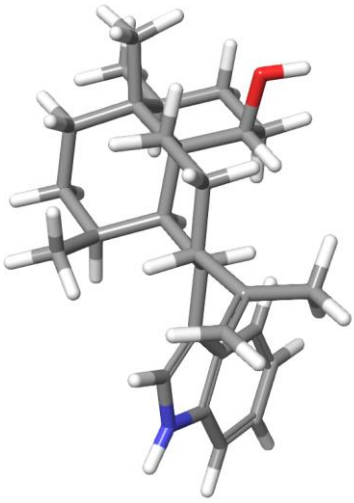


**Figure 3.17. 3D structure of epi-10,23-dihydro-24,25-dehydroaflavine with the best correlation to the NMR data.**

**Table 3.19. Comparison of the distances from NOE intensities and from 3D structures using a single DFT optimized structure (B3LYP/6-31g(d)) or multiple structures from MM ($\tau_m$ = 150 ms). All distances in Å. [a]Reference.**

| | | Single | | | Mult | | |
|---|---|---|---|---|---|---|---|
| H1 | H2 | Exp. Dist. | Calc. Dist. | % | Exp. Dist. | Calc. Dist. | % |
| 1 | 2 | 2.64 | 2.50 | 5.6 | 2.62 | 2.48 | 5.5 |
| 1 | 5 | 3.03 | 2.83 | 7.1 | 3.01 | 2.81 | 6.7 |
| 5 | 18 | 2.72 | 2.40 | 13.3 | 2.70 | 2.46 | 8.8 |
| 5 | 11 | 2.41 | 2.37 | 1.7 | 2.40 | 2.62 | 9.3 |
| 5 | 10 | 2.54 | 2.67 | 4.9 | 2.53 | 2.56 | 1.5 |
| 2 | 27 | 2.89 | 2.96 | 2.4 | 2.88 | 2.99 | 3.5 |
| 2 | 23 | 2.58 | 2.46 | 4.9 | 2.56 | 2.56 | 0.1 |
| 2 | 25 | 3.01 | 2.96 | 1.7 | 3.00 | 3.04 | 1.3 |
| 25 | 23 | 2.39 | 2.31 | 3.5 | 2.37 | 2.39 | 0.9 |
| 25 | 10 | 2.78 | 4.58 | 39.3 | 2.77 | 2.82 | 2.0 |
| 25 | 25’ | 1.85 | 1.85 | 0.0 | 1.84 | 1.85 | 0.6 |
| 19 | 26 | 3.57 | 3.53 | 1.0 | 3.56 | 3.64 | 2.3 |
| 19 | 18’ | 2.33 | 2.43 | 3.3 | 2.33 | 2.45 | 5.0 |
| 19 | 22 | 2.02 | 2.14 | 5.6 | 2.01 | 2.23 | 10.7 |
| 19 | 18 | 2.32 | 2.41 | 3.7 | 2.31 | 2.44 | 5.5 |
| 19 | 11 | 3.37 | 2.83 | 19.1 | 3.35 | 2.95 | 12 |
| 19 | 10 | 1.95 | 1.88 | 3.7 | 1.93 | 2.01 | 3.8 |
| 30 | 29 | 3.45 | 3.39 | 1.8 | 3.44 | 3.65 | 6.0 |
| 30 | 18’ | 2.34 | 3.32 | 25.3 | 2.33 | 2.41 | 3.5 |
| 30 | 22 | 2.68 | 2.25 | 19.1 | 2.66 | 2.70 | 1.3 |
| 30 | 21 | 2.65 | 2.06 | 28.6 | 2.64 | 2.54 | 3.6 |
| 10 | 26 | 2.28 | 2.33 | 2.1 | 2.72 | 2.86 | 5.4 |
| 10 | 11 | 2.24 | 2.32 | 3.4 | 2.23 | 2.33 | 4.7 |
| 23 | 27 | 2.06 | 2.05 | 0.5 | 2.40 | 2.49 | 3.7 |
| 11 | 12 | 2.12 | 2.24 | 5.4 | 2.18 | 2.28 | 4.5 |
| 11 | 13 | 2.70 | 2.61 | 3.4 | 2.68 | 2.67 | 0.3 |
| 11 | 16 | 1.86 | 2.21 | 15.8 | 1.85 | 2.00 | 7.7 |
| 16 | 28 | 2.68 | 2.84 | 5.9 | 2.67 | 2.60 | 2.6 |
| 21 | 29 | 2.46 | 2.61 | 6.0 | 2.45 | 2.64 | 7.6 |
| 16 | 17’ | 2.10 | 2.44 | 13.9 | 2.09 | 2.16 | 3.4 |
| 16 | 13 | 1.93 | 2.02 | 4.5 | 1.91 | 2.09 | 9.0 |
| 22 | 22' | 1.75 | 1.74 | 0.6 | 1.74 | 1.74 | 0.0 |
| 17 | 29 | 2.77 | 2.79 | 0.9 | 2.76 | 2.84 | 3.0 |
| 17 | 17’ | 1.74 | 1.75 | 0.6 | 1.71 | 1.75 | 2.3 |
| 21’ | 27 | 2.58 | 2.52 | 2.4 | 2.57 | 2.62 | 1.7 |
| 13 | 13’ | 1.75[a] | 1.75 | 0.0 | 1.74[a] | 1.74 | 0.0 |
| 17’ | 28 | 2.82 | 2.95 | 4.5 | 2.81 | 2.87 | 2.1 |
| 14’ | 28 | 2.54 | 2.60 | 2.3 | 2.53 | 2.70 | 6.8 |
| 25’ | 26 | 2.85 | 2.92 | 2.5 | 2.84 | 2.88 | 1.2 |
| | MAE | | | 6.9 | | | 4.1 |

**Table 3.20 Comparison of experimental *J*-coupling constants and calculated constants by the HLA equation or DFT calculation from a single DFT optimized 3D structure, in Hz.**

| H1 | H2 | Exp. | HLA | DFT |
|---|---|---|---|---|
| 10 | 23 | 13.3 | 12.7 | 12.4 |
| 10 | 11 | 5.0 | 4.3 | 5.1 |
| 21' | 22 | 13.5 | 12.6 | 11.9 |
| 22 | 23 | 13.5 | 11.9 | 13.7 |
| 16 | 17 | 11.3 | 12.5 | 13.7 |
| 17 | 18 | 11.3 | 13.4 | 14.1 |
| 21' | 22' | 3.4 | 4.8 | 4.4 |
| 21 | 22 | 4.3 | 4.5 | 4.4 |
| 22' | 23 | 5.6 | 4.8 | 5.3 |

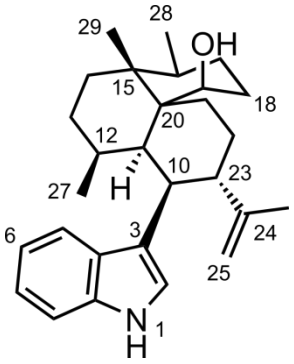



**Figure 3.18. Numbered structure of epi-10,23-dihydro-24,25-dehydroaflavine.**

## 3.3 Synthetic peptides

*The experimental parts of this section are based on the articles below. More space is used discussing the results from NMR spectroscopy than the various biological roles and otherwise exiting potential of the included compounds, which will only be briefly touched upon. For a more in depth description the reader is referred to the articles below and the articles referred to in the text.*

Appendix A11: Methyl Effect in Azumamides Provides Insight Into Histone Deacetylase Inhibition by Macrocycles
Appendix A11: Specific Electrostatic Molecular Recognition in Water

### 3.3.1 Azumamides

The azumamides are natural products, originally isolated from the marine sponge *Mycale izuensis*.[90] The compounds are cyclic non-ribosomal peptides (NRPs), illustrated in Figure 3.19, which were shown to inhibit histone deacetylase (HDAC) activity.[90]

| | $R_1$ | $R_2$ | $R_3$ |
|---|---|---|---|
| **A:** | $NH_2$ | H | Me |
| **B:** | $NH_2$ | OH | Me |
| **C:** | OH | OH | Me |
| **D:** | $NH_2$ | H | H |
| **E:** | OH | H | Me |

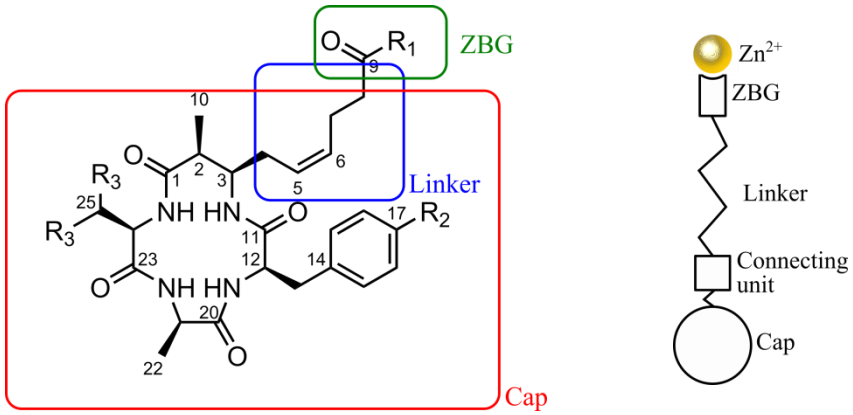



**Figure 3.19. The structure of the natural azumamides A-E with indicated pharmacophore, and a schematic pharmacophore of HDAC inhibitors.[90,91]**

HDACs are a group of enzymes with the common trait that they deacetylate the *N*-ε-position of lysine residues at the *N*-terminal tail of histones.[91,92] In groups of eight, histones proteins form the nucleosome; the repeating unit in chromatin, which is the highly ordered structures that store DNA in eukaryotic cells.[91,92] Gene-transcription is regulated by interactions between the *N*-terminal end of the histones and DNA, and the interactions are modulated by e.g. the degree of histone acetylation, methylation or phophorylation.[91,92] Histone acetylation will therefore affect gene transcription as the degree of histone acetylation correlates positively with transcription,[92] and a decrease in acetylation has been linked to cancer due to a lower expression of regulatory genes.[91] Compounds that inhibit HDAC activity may thus be used as anti-cancer agents, spurring the interest in understanding and improving the activity of the compounds. Humans have 18 HDACs, of which 11 are $Zn^{2+}$ dependent and were the focus of the study.

The mode of action of most HDAC inhibitors is by interacting with $Zn^{2+}$ in the binding site of the HDAC enzymes, effectively blocking the active site. HDAC inhibitors are structurally quite different compounds but generally have a pharmacophore as indicated in Figure 3.19.[91]

The group of Professor Christian A. Olsen (CAO), then DTU now University of Copenhagen, were interested in investigating analogues of the natural azumamides, synthesized in lab from total

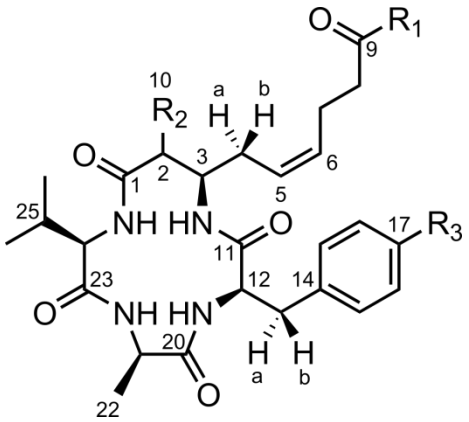
synthesis, to develop novel anti-cancer compounds, and it was in that regard desired to correlate structural features to HDAC inhibition. Thus an NMR study of natural and unnatural analogues of azumamides was conducted in collaboration with Dr. Alex Maolanon and Dr. Jesper Villadsen both from the group of CAO, who synthesized the investigated molecules, and Dr. Niels Christensen from the group of Associate Professor Peter Fristrup, who performed the calculations and theoretical docking studies. For more information regarding the comprehensive synthetic work, activity testing and docking results the reader is referred to the article.[93]

**Prior work**

A lot of work has been put towards total synthesis and HDAC activity of the native azumamides as well as structural analogues thereof.[94,95] This, often synthetic or biological work, is mostly outside the scope of this thesis and it was chosen to keep the focus relevant to the presented work – namely 3D structural information from NMR spectroscopy. The 3D structures of azumamide E[94,95] and an unnatural analogue 2*S*,3*R* azumamide E[95] have previously been published from constrained MD calculations. The comparable structures of azumamide E are alike; the amide protons NH-1,3 and 4 points upwards from the plane of the ring while NH-2 points slightly downwards and the phenyl ring is superimposing the ring akin to a lid. These structures are used for comparison in the following sections.

**Investigated structures**

Three structures were chosen for NMR investigation; the natural azumamide A (azu A) and two unnatural analogues desmethyl-azumamide C and epimethyl-azumamide E (des D and epi E). The structures are given in Figure 3.20.




| | $R_1$ | $R_2$ | $R_3$ |
|---|---|---|---|
| **Azu A:** | $NH_2$ | (*S*)-Me | H |
| **Des C:** | OH | H | OH |
| **Epi E:** | OH | (*R*)-Me | H |

**Figure 3.20. Structure of the investigated structural analogues azumamide A (azu A), des-azumamide C (des C) and epi-azumamide E (epi E). Stereochemistry, including the diastereotopic protons important for the 3D structural investigation, is indicated.**

The structural analogues exhibited *in vitro* activity in the order azumamides > des-methyl azumamides >> epi-methyl azumamides.[93] It was proposed that conformational differences in the ring conformation could be the reason for the differences in activity. The slight differences in the groups $R_1$-$R_3$ (Figure 3.20) were deemed of lesser importance for the ring conformation, and the general conformational spaces occupied by the structures. Unfortunately the *in vitro* activity was not retained in *in vivo* studies of a human Burkitt's lymphoma cancer cell line.

As the interesting properties of the azumamides are all occurring in an aqueous environment, it would be beneficial to obtain data in water. Attempts to dissolve the compounds in water and in a water/DMSO-$d_6$ mixture with up to 15 % DMSO-$d_6$ resulted in very low *S*/*N* ratios, and spectra acquired in pure DMSO-$d_6$ were used for further investigations. It is acknowledged that this may lead to errors in the structure compared to that of a biological environment, but it could not be averted.

**Assignment of diastereotopic protons 4 and 13**

The diastereotopic protons at C-4 and C-13 were not easily assigned due to possible free rotation around the bonds C3-C4 and C12-C13. It was realized from NOE-derived distances and *J*-coupling constants, that free rotation was generally not present, due to differences in NOE intensities and the size of *J*-coupling constants to the diastereotopic protons, allowing assignment of the protons. Only the rotation around C3-C4 of the des-methyl variant was observed to be unrestricted. The protons were thus assigned by minimizing the error of NOEs and *J*-coupling constants. This might not lead to the right answer, but without an assignment a lot of data and information was lost, and both data types suggest the given assignments unambiguously, making it the most viable option.

The assignment results in structures that fit the NMR data displaying the aromatic ring of Phe or Tyr outside the cylinder of the ring, in contrast to previously published structures, which is in good correlation with docking studies, *vide infra*.

**Constrained versus unconstrained calculations**

Constrained MD using experimental observables such as relative NOE intensities and *J*-coupling constants has been a favored method of obtaining 3D structures.[94,95] While the approach has validity for macromolecules, where the amount of flexibility is beyond what even modern computational chemistry can reasonably be expected to handle, for relatively ordered systems a lot of information will inevitable be lost. This is in part due to the constrains being applied as distances and dihedral angles, while the observables are in fact NOEs and *J*-coupling constants, which will lead to a faulty averaging of the data, see equations (3.2)-(3.5). Here it should be noted that not all structural simulation programs are capable of performing an averaging routine; it is obvious that information is lost if averaging is disregarded. The distinction in (3.2)-(3.5) may seem of less importance but will lead to large error in conformational populations.

$$\eta_{obs,i} = k\, r_i^{-6} \quad \text{for} \quad i = 1 \tag{3.2}$$

$$\eta_{obs,i} = k\, \langle r_i^{-6} \rangle \quad \text{for} \quad i > 1 \quad \text{not} \quad \eta_{obs,i} = k\, \langle r_i \rangle^{-6} \tag{3.3}$$

$$J_{obs,i} = f(\theta_i) \quad \text{for} \quad i = 1 \tag{3.4}$$

$$J_{obs,i} = \langle f(\theta_i) \rangle \quad \text{for} \quad i > 1 \quad \text{not} \quad J_{obs,i} = f(\langle \theta_i \rangle) \tag{3.5}$$

The other possibility is to use unconstrained simulations and structural averaging of conformers. This is an approach used earlier in this thesis that is gaining popularity and being applied in a slightly altered form also in macromolecular NMR.[96] To achieve this, a couple of requirements are apparent. First the simulations need to cover the structural space that the actual *in vitro* compounds occupy. But if too much space is covered it may hamper the fitting of the experimental data to the

3D structures. This means that while an amount of conformational flexibility is a good thing, too much flexibility may lead to unsolvable systems. Also one has to accept a couple of approximations; that the correlation time is independent of conformation and that the averaging motions are of a timescale that allows the usage of $r^{-6}$ averaging. For the azumamides, due to an at least somewhat restricted conformational space due to the cyclic system, it should be possible to use unconstrained simulations and increase the coverage of the conformational space and thus obtainable molecular knowledge. For comparison both strategies were tried.

**Constrained structure**

Constraints in the form of NOE distances (Table 3.21) were applied to a structure of azumamide A, and the structure was optimized, depicted in Figure 3.21. This structure was in good correlation to the published structures of azumamide E from constrained optimizations, with the exception of the orientation of the aromatic ring as described earlier, and the extension of the linker chain, which was generally less ordered, and was not considered beyond the double-bond.[94,95]

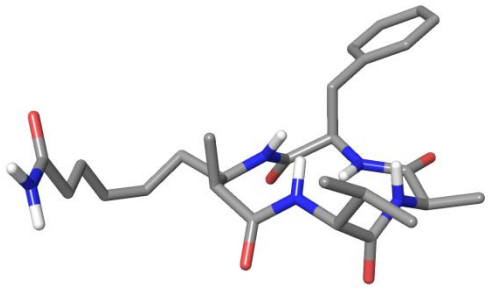


**Figure 3.21. Representational 3D structure from the constrained optimization of azumamide A. Structures of azumamide E are available from the literature for comparison.[94,95]**

When investigating the NOE data, the constrained structure was in clear violation of multiple distances around the NH-2 site, which indicated that this structure was probably not that good a representation of the data, see Appendix A4.

**Unconstrained structures**

Unconstrained simulations were conducted in water and DMSO for the three azumamide analogues. The resulting conformational spaces were overlapping and the structures from water were used. Most distances and *J*-coupling constants were comparable between the three analogues, though some differences were observed. When fitting the NMR data to multiple structures this was translated to a mostly identical conformational space between the analogues. The correlations of the data to the simulated structures were good with MAE% of 3.1 %, 3.5 % and 3.7 % for Azu A, Des C and Epi E respectively. Also the *J*-coupling constants exhibited a good correlation between experimental and back-calculated values. The structures from unconstrained simulations were thus deemed to be better representations of the NMR data than the structures from constrained simulations and further conclusion were based on these. The experimental distances and back-calculated values, obtained by optimizing the correlation of 2,500 structures from molecular dynamics iteratively, are shown in Table 3.21.

**Table 3.21. Experimental and back-calculated distances in Å, including differences in %, and *J*-coupling constants in Hz for the azumamide analogues ($\tau_m$ = 150 ms). [a]Reference. Based on 2,500 structures each, from MD simulations.**

**Azumamide A**

| H1 | H2 | Exp. | Calc. | % |
|---|---|---|---|---|
| N1 | N2 | 2.13 | 2.15 | 1.3 |
| N1 | N4 | 3.11 | 3.11 | 0.0 |
| N1 | 3 | 2.62 | 2.66 | 1.4 |
| N1 | 4a | 3.12 | 3.12 | 0.0 |
| N1 | 4b | 3.06 | 3.06 | 0.0 |
| N1 | 10 | 2.75 | 2.80 | 1.8 |
| N1 | 13a | 2.49 | 2.49 | 0.0 |
| N1 | 13b | 2.42 | 2.44 | 0.7 |
| N2 | 13a | 2.64 | 2.64 | 0.0 |
| N2 | 13b | 2.89 | 2.89 | 0.0 |
| N2 | 21 | 2.61 | 2.86 | 8.8 |
| N2 | 22 | 2.83 | 3.15 | 9.9 |
| N3 | N4 | 2.38 | 2.35 | 1.0 |
| N3 | 22 | 2.80 | 2.97 | 5.6 |
| N3 | 21 | 2.71 | 2.87 | 5.5 |
| N3 | 25 | 2.45 | 2.45 | 0.0 |
| N3 | 26/27 | 3.13 | 3.13 | 0.0 |
| N4 | 2 | 3.24 | 3.33 | 2.5 |
| N4 | 10 | 2.67 | 2.49 | 7.3 |
| N4 | 24 | 2.96 | 2.97 | 0.4 |
| N4 | 25 | 3.07 | 2.68 | 14.7 |
| N4 | 26/27 | 3.17 | 2.97 | 7.0 |
| 2 | 10 | 2.59 | 2.68 | 3.2 |
| 2 | 3 | 2.13 | 2.39 | 10.7 |
| 2 | 4a | 2.76 | 2.76 | 0.0 |
| 4b | 4a | 1.82 | 1.82 | 0.0 |
| 4a | 10 | 2.71 | 2.97 | 8.7 |
| 4b | 10 | 3.52 | 3.51 | 0.3 |
| 12 | 15 | 2.65 | 2.65 | 0.0 |
| 13a | 15 | 2.50 | 2.44 | 2.5 |
| 13b | 15 | 2.47 | 2.47 | 0.0 |
| 21 | 22 | 2.77 | 2.77 | 0.1 |
| 24 | 25 | 2.98 | 2.85 | 4.7 |
| 24 | 26/27 | 2.62 | 2.81 | 6.9 |
| 25 | 26/27 | 2.50 | 2.43 | 2.7 |
| MAE% | | | | 3.1 |

***J*-coupling constants**

| H1 | H2 | Exp. | Calc. | Diff |
|---|---|---|---|---|
| N1 | 3 | 8.0 | 8.0 | 0.0 |
| N2 | 12 | 9.1 | 8.9 | 0.2 |
| N3 | 21 | 8.7 | 8.2 | 0.5 |
| N4 | 24 | 8.3 | 8.3 | 0.0 |
| 2 | 3 | 3.8 | 3.8 | 0.0 |
| 3 | 4a | 8.1 | 8.1 | 0.0 |
| 3 | 4b | 6.4 | 6.4 | 0.0 |
| 12 | 13a | 9.4 | 8.8 | 0.6 |
| 12 | 13b | 6.7 | 6.7 | 0.0 |
| 24 | 25 | 10.1 | 10.1 | 0.0 |

**Des-azumamide C**

| H1 | H2 | Exp. | Calc. | % |
|---|---|---|---|---|
| N1 | N3 | 2.79 | 2.80 | 0.4 |
| N1 | 2b | 2.29 | 2.49 | 7.8 |
| N1 | 3 | 2.73 | 2.70 | 1.0 |
| N1 | 4a/4b | 3.11 | 3.00 | 3.6 |
| N1 | 12 | 2.98 | 3.11 | 4.5 |
| N1 | 13a | 2.79 | 2.72 | 2.3 |
| N1 | 13b | 2.99 | 3.00 | 0.6 |
| N1 | N2 | 1.99 | 1.91 | 4.4 |
| N2 | N3 | 2.21 | 2.20 | 0.0 |
| N2 | 13a | 2.56 | 2.59 | 1.2 |
| N2 | 13b | 2.89 | 2.90 | 0.3 |
| N2 | 21 | 3.38 | 3.55 | 4.8 |
| N2 | 22 | 2.83 | 2.87 | 1.3 |
| N3 | N4 | 2.30 | 2.37 | 2.7 |
| N3 | 22 | 2.94 | 2.96 | 0.6 |
| N3 | 24 | 3.36 | 3.45 | 2.6 |
| N3 | 25 | 2.88 | 2.99 | 3.6 |
| N3 | 26/27 | 3.78 | 3.79 | 0.2 |
| N4 | 2b | 2.02 | 2.18 | 7.1 |
| N4 | 25 | 2.59 | 2.56 | 1.1 |
| N4 | 26/27 | 3.07 | 3.07 | 0.0 |
| 2b | 2a | 1.79[a] | 1.79 | 0.0 |
| 3 | 2a/4a/4b | 1.99 | 2.07 | 3.8 |
| 5 | 3 | 2.96 | 2.94 | 0.5 |
| 6 | 7 | 2.45 | 2.45 | 0.2 |
| 12 | 15 | 3.02 | 2.86 | 5.7 |
| 13b | 13a | 2.08 | 1.70 | 22.1 |
| 13a | 15 | 2.72 | 2.44 | 11.3 |
| 13b | 15 | 2.66 | 2.47 | 8.0 |
| 15 | 22 | 4.31 | 4.25 | 1.4 |
| 21 | 22 | 2.60 | 2.62 | 0.5 |
| 24 | 25 | 2.44 | 2.56 | 4.8 |
| 24 | 26/27 | 3.02 | 3.19 | 5.4 |
| 25 | 26/27 | 2.74 | 2.59 | 5.7 |
| MAE% | | | | 3.5 |

***J*-coupling constants**

| H1 | H2 | Exp. | Calc. | Diff |
|---|---|---|---|---|
| N1 | 3 | 7.3 | 7.4 | 0.1 |
| N2 | 12 | 9.6 | 9.1 | 0.4 |
| N3 | 21 | 9.7 | 9.0 | 0.7 |
| N4 | 24 | 6.8 | 6.8 | 0.0 |
| 2b | 3 | 12.4 | 12.4 | 0.0 |
| 2a | 3 | 3.6 | 3.6 | 0.0 |
| 3 | 4a | 6.5 | 6.5 | 0.0 |
| 3 | 4b | 6.5 | 6.6 | 0.1 |
| 12 | 13a | 9.3 | 9.3 | 0.0 |
| 12 | 13b | 5.8 | 5.8 | 0.0 |
| 24 | 25 | 6.7 | 6.8 | 0.1 |

**Epi-azumamide E**

| H1 | H2 | Exp. | Calc. | % |
|---|---|---|---|---|
| N1 | N2 | 1.88 | 1.91 | 1.1 |
| N1 | 2 | 2.50 | 2.49 | 0.3 |
| N4 | 2 | 2.15 | 2.12 | 1.1 |
| N2 | 13a | 2.92 | 2.92 | 0.1 |
| N2 | 22 | 3.12 | 3.13 | 0.3 |
| N3 | N4 | 2.39 | 2.29 | 4.2 |
| N4 | 25 | 2.98 | 2.99 | 0.1 |
| N4 | 26/27 | 2.93 | 2.75 | 6.6 |
| 2 | 3[a] | 3.00 | 3.00 | 0.0 |
| 2 | 10 | 2.82 | 2.74 | 3.1 |
| 3 | 4a | 2.34 | 2.35 | 0.5 |
| 3 | 4b | 2.53 | 2.62 | 3.3 |
| 3 | 10 | 3.05 | 3.07 | 0.6 |
| 4b | 10 | 2.86 | 2.90 | 1.4 |
| 6 | 7 | 2.16 | 2.32 | 6.9 |
| 12 | 15 | 2.92 | 2.80 | 4.2 |
| 21 | 22 | 2.65 | 2.78 | 4.6 |
| 24 | 25 | 2.57 | 2.66 | 3.4 |
| 24 | 26/27 | 2.71 | 2.87 | 5.4 |
| 25 | 26/27 | 2.59 | 2.43 | 6.5 |
| MAE% | | | | 2.9 |

***J*-coupling constants**

| H1 | H2 | Exp. | Calc. | Diff |
|---|---|---|---|---|
| N1 | 3 | 8.7 | 9.5 | 0.8 |
| N2 | 12 | 8.9 | 10.9 | 2 |
| N3 | 21 | 9.1 | 11.6 | 2.5 |
| N4 | 24 | 8.5 | 9.4 | 0.9 |
| 2 | 3 | 12.4 | 13.0 | 0.6 |
| 3 | 4a | 4.6 | 4.6 | 0.0 |
| 3 | 4b | 5.6 | 5.6 | 0.0 |
| 12 | 13a | 10.0 | 10.5 | 0.5 |
| 12 | 13b | 5.2 | 5.4 | 0.2 |

**Structural differences**

Using unconstrained simulations, the difference in the ring conformation between the three compounds was diminishing. It would seem as if the natural compound had more flexibility in the ring, but the majority of the conformers which fitted the NMR data had NH-2 pointing upwards, in line with the synthetic structures. The best correlation to the data had NH-2 pointing upwards approximately 90 % of the time, and downwards 10 %. This is proposed to be translated to the intermediate position in the constrained calculations. A similar distribution was also seen for Des C. The multiple conformer approach led to a much better fit for the ring, than the constrained simulations – still with errors around the NH-2 though, see Table 3.21.

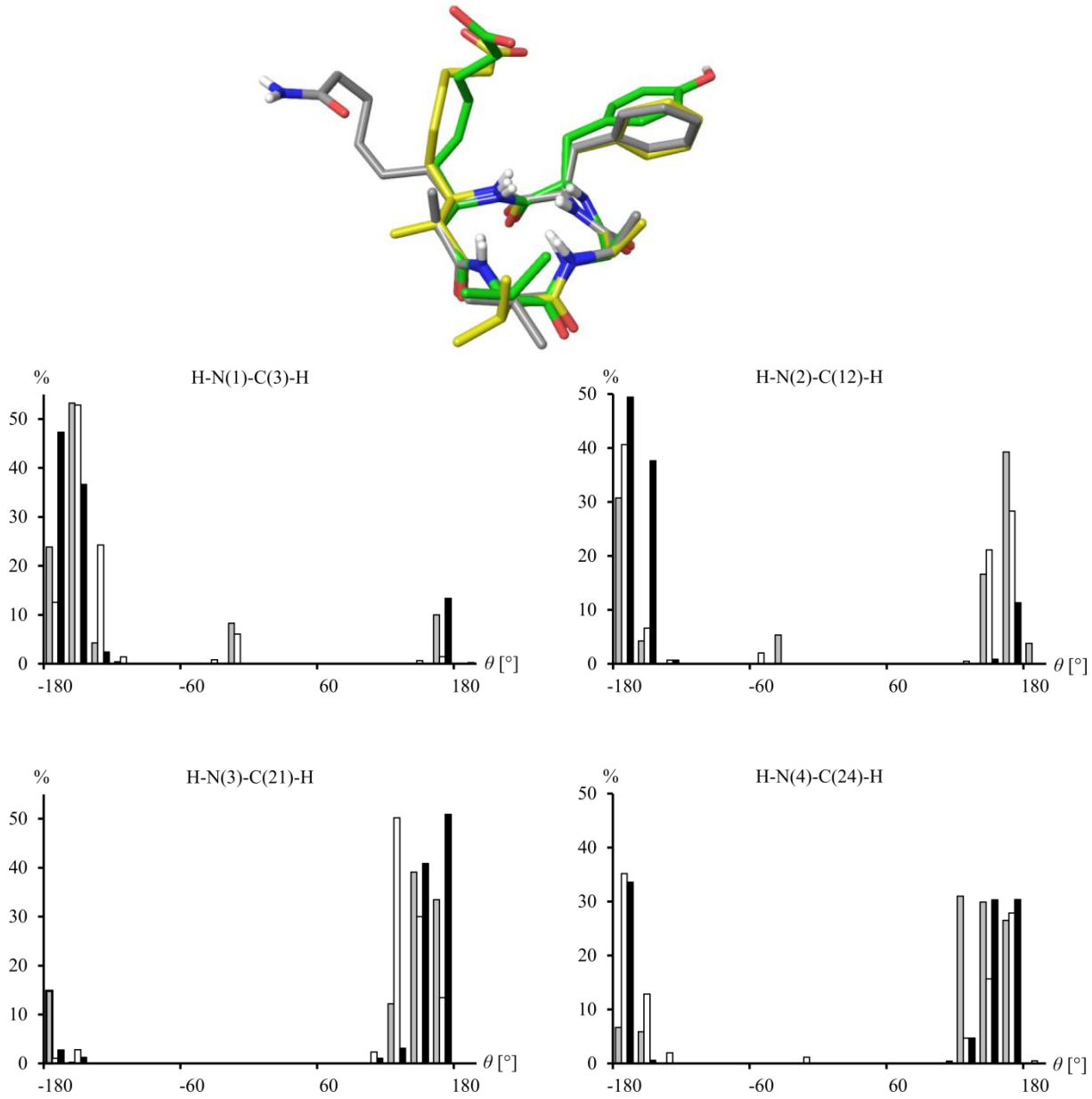



**Figure 3.22. Top: The best fit structure of the three analogues overlain; grey: Azu A, green: Des C, yellow: Epi E. Bottom: NH-$C_\alpha$H dihedral angles of the conformers used in the NOE and J-coupling constant fit of azu A (grey), des C (white) and epi E (black).**

The ring conformation was thus essentially unaffected by the inversion or removal of the methyl group, in good correlation to the NMR data and simulations. This led to the search for, and identification of, other differences between the three analogues.

**Side-chain orientation**

The side-chain orientation was the only major difference between the different compounds. While the distances between the side-chains and the ring generally differed more than the ring proton distances, it was only translated into major differences in the orientation for the linker side-chain, which in turn is arguably the most important side-chain considering the involvement in reaching the binding site.[91]

The side-chain of the linker was thus the only clear difference in NMR observables between the three compounds. This was also apparent from the simulations, but is more pronounced, and experimentally confirmed, in the NMR data. The populations of the orientations for the different structural analogues are found in Figure 3.23.

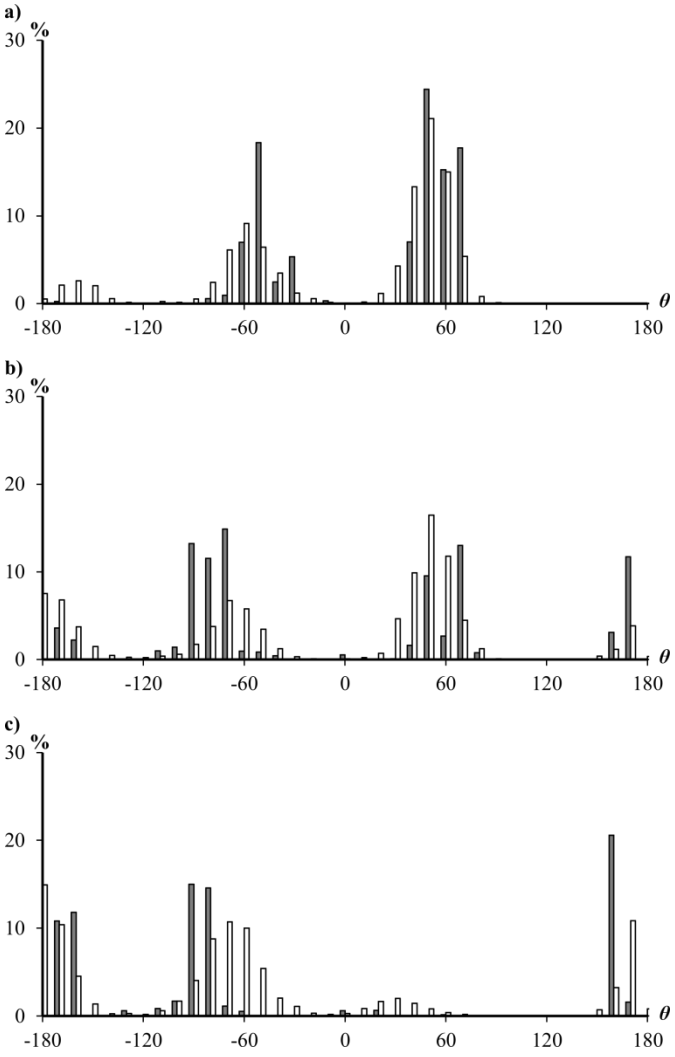
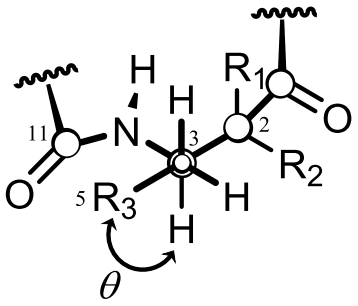



**Figure 3.23. The population in % of the orientations of the side-chain at C-3, illustrated by the dihedral angle H3-C3-C4-C5 ($\theta$), for the natural azu A (a), and the synthetic des C (b) and epi E (c).**

**Grey: Population that fit the NMR data the best.**

**White: Full simulation (2,500 structures).**




**Azu A:$R_1$=-Me $R_2$=-H $R_3$=-CH=CHCH$_2$CONH$_2$**
**Des C:$R_1$=-H $R_2$=-H $R_3$=-CH=CHCH$_2$CONH$_2$**
**Epi E:$R_1$=-H $R_2$=-Me $R_3$=-CH=CHCH$_2$CONH$_2$**

The natural compound, which is the most active, had an orientation of the linker side-chain centered on $\theta$ = ±60°. The des-methyl analogue displayed an orientation shared between the three sites in a three jump model, while the epi-methyl, the least active analogue, displayed an orientation centered on $\theta$ = -80° and 180°. It may thus be proposed that favorable activity would be obtained when the orientation of the linker was at $\theta$ = +60° in the binding pocket, as this would correlate well to the activity data. This result was compared to results from the theoretical docking studies.

**Docking results**

The chosen HDAC to be used in the theoretical docking was HDA3, as a crystal structure of a co-repressor complex was available to dock into. The resulting azumamide structures had all amide protons directed to the same face and toward the carboxylate side-chain of Asp93, in agreement with solution data. The orientation of the aromatic side-chain also correlated well to the solution conformations. The linker side-chain had a dihedral angle of either 180° or 60°, except for azumamide E where a dihedral of 110° was prevalent. It was thus not possible to correlate the preferable dihedral angle from NMR to the docking studies, and since no docked crystal structure is available, a possible connection must be regarded as a possibility and supported speculations. It is of course not necessary for the bound compound to be of the same conformation as in the free liquid state, but it would lead to a potentially lower energy penalty upon binding. From theoretical calculations of the binding interactions, it was determined that the differences in activity should probably be contributed to a large methyl effect from interactions to Phe200 of the enzyme.

**Minor conformer**

It was apparent from the spectra that a minor conformational average was present in solution that averaged slowly on an NMR time scale from multiple resonances in the spectra. The data was thus catered for as discussed earlier for the natural compound CM-A. In contrast to CM-A, where the minor conformer was not investigated further, it was speculated that the minor conformer could be a *cis*-amide arrangement of one of the amide bonds in the ring. This is supported experimentally by an observed strong NOE between H-12 and H-21, leading to the cis-amide bond being between N-2 and C-20, the most flexible amide bond in the major conformational average. The 3D structures of these conformations were not determined due to weak NOE intensities. The differences between the ratios of the conformations were not significant between the different analogues and could thus not be used to explain the differences in the activity.

**Conclusion**

Solution state structures of natural and analogues of representative azumamides were investigated by NOE and *J*-coupling constants leading to knowledge of the structural space inhabited by the molecules. The ring structures of all were more or less equal with few differences between analogues. The orientation of the $\beta^3$-chain differed from natural azumamide to the unnatural structures, but this could not be correlated to theoretical docked structures. Still the knowledge of the structure was increased by applying a non-constrained methodology compared to previous publications, where a constrained simulation approach was utilized.

### 3.3.2 Molecular recognition

Molecular recognition is defined as a specific non-covalent interaction of two or more molecules and is an essential part of many highly important biological functions, from protein folding, enzyme reactivity and our immune system to the structural organization of cells.[97–99] Due to the widespread dependence on molecular recognition in human biology, gaining further understanding on the involved processes is important in the development of drugs that e.g. interfere with specific protein-protein interactions.[100]

PhD student Ming Li of the group of Professor Morten Meldal (MM), University of Copenhagen, developed a methodology of identifying peptide partners which bound with high specificity from on bead pairing of a large library of approximately 78,125 compounds. The bead-based screening protocol, developed in the lab of MM, allows for a simultaneous screening of up to $10^{12}$ peptide-peptide interactions on the biocompatible poly(acryloyl-bis-(aminopropyl)polyethylene glycol) (PEGA) resins.[101,102] The screening was setup as to identify peptides which bound from electrostatic interactions, which was investigated using competitive binding condition, where salt concentrations were shown to influence the binding. Lastly the binding was evaluated by fluorescence microscopy using a custom-made flow-cell.[103] For more information regarding the comprehensive work of developing and utilizing the screening method, the reader is referred to the article.[104] Multiple binding partners were identified and a condensed conclusion of their findings is found in Table 3.22 which displays the best binding partners for the peptide Target 20.

**Table 3.22. Binding constants of various peptides (ligands) with Target 20.**

| Ligand # | Structure | $k_d$ [µM] |
|---|---|---|
| 7 | EDYEVEEG | 0.658±0.029 |
| 17 | EDDWDDG | 3.37±0.25 |
| 18 | EDYEVDDG | 6.95±1.49 |
| 19 | EDYEWEEG | 0.661±0.059 |

The structural build of the best binding partners were all conserved, as amino acids with negatively charged side-chains in both ends and a single hydrophobic amino acid in the middle of the peptide, suggesting an electrostatic interaction. This is in contrast to many previous reports on molecular recognition, where hydrophobic interactions in water is usually the principal driving force towards binding.[100,105–108] The *C*-terminal was kept a glycine due to easier coupling and cleavage of the peptide and resin.

**NMR study**

From the screening two peptides were chosen for structural elucidation; Target 20 (T20) and Ligand 7 (L7), shown in Figure 3.24, which had shown a high degree of specific binding. T20 consists of ten amino acids while L7 consists of eight. The 3D structures of these compounds were investigated using distances from NOE data.

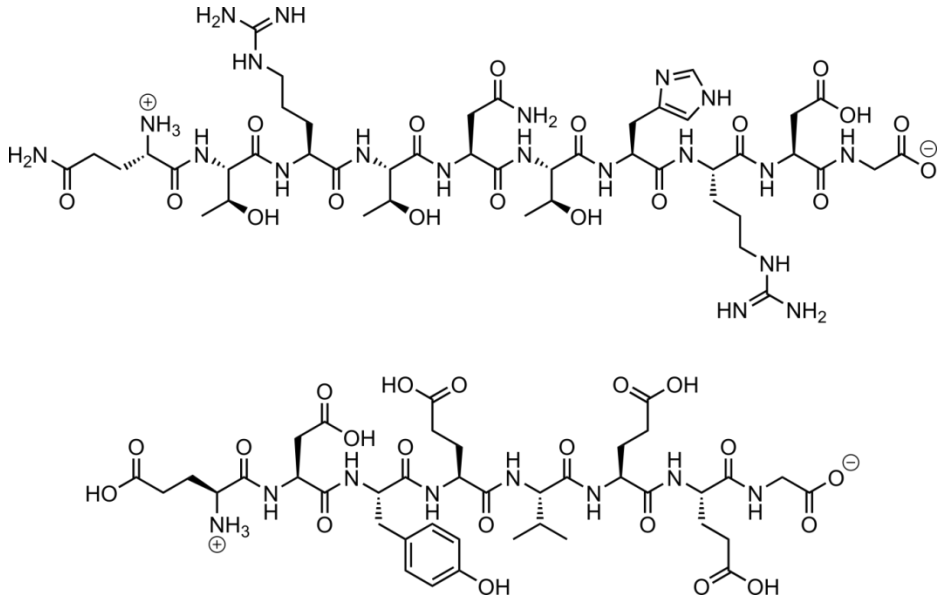


**Figure 3.24. Top: The structure of Target 20 (T20). Bottom: The structure of Ligand 7 (L7).**

The two peptides display a specific binding interaction with a $k_D$ of 0.7 μM. NMR data were obtained for the two molecules in isolation and of a mixture of the two. The assignments of the compound resonances are found in Appendix A4. Upon mixing the chemical shifts were almost unchanged, especially for T20, though some changes were observed in the form of differentiation of chemical shifts of overlapped peaks for L7, see Figure 3.25. In contrast to the determined low binding constant, no intermolecular NOE correlations were observed. Some changes in distances of the individual peptides were observed upon mixing, but the changes were in most cases minor.

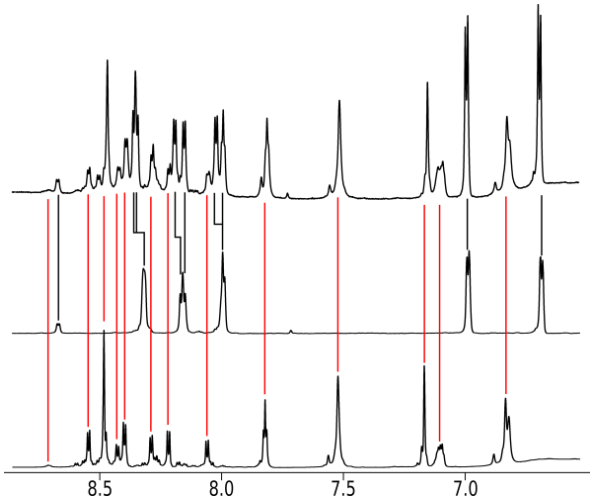



**Figure 3.25. Change in chemical shifts in the amide region upon mixing T20 and L7 at pH 6. The top spectrum is the mix, the middle spectrum is L7 and bottom spectrum is T20.**

### Differences in the rotational correlation time

An indication of interaction was an apparent increase in the rotational correlation time ($\tau_c$) for both peptides upon mixing. This would be in correlation with an interaction so that both peptides behave as bigger molecules in solution. This may be shown utilizing the work of Macura *et al.* (equations (2.3) and (2.11)), making (3.6) valid for short mixing times.[17]

$$\frac{\eta_{IS}(t_m)}{½[\eta_I(t_m)+\eta_S(t_m)]} = -(W_{2IS}-W_{0IS})t_m = \sigma_{IS}t_m$$
$$\sigma_{IS} = \left(\frac{\mu_0}{4\pi}\right)^2 \frac{\hbar^2\gamma^2}{10}\left(\tau_c - \frac{6\tau_c}{1+4\omega^2{\tau_c}^2}\right) r_{IS}^{-6} \quad \textbf{(3.6)}$$

**Equation (3.6) is a revisit from Section 2.1 and is valid for homonuclear NOEs, where $\omega$ is approximately 5·109 rad/s at 800 MHz. As seen, the value of $\sigma_{IS}$ is dependent on the distance $r$ and $\tau_c$. For e.g. methylene proton pairs where the distance is known, the correlation time can be calculated. The shift in $\tau_c$ was immediately visible in the NOESY spectra as the cross-peak to diagonal-peak intensity was significantly higher for the mixture compared to the single peptides, while the summed concentration was equal and thus lower for the individual peptide. Since the NOEs were in-phase with the diagonal and thus higher than the crossover correlation time, an increase in the size of the peptide would increase the NOE intensities as given in equation (3.6) and thereby correlated to a longer $\tau_c$ (see Figure 2.4 p. 6). It was observed that the $\sigma$ for proton pairs generally increased, especially for T20, which translates to either a majority of the distances being shorter or a difference in $\tau_c$; the latter being the most probable. Both scenarios indicated an interaction between the peptides though. For the methylene proton pairs $\tau_c$ could be calculated by assuming a fixed distance of 1.78 Å for methylene protons. Comparisons are found in Table 3.23 and**

Table 3.24 for the Target, Ligand and the mix of the two.

**Table 3.23. Rotational correlation times ($\tau_c$) and differences in cross-relaxation rate constants for protons in of T20, in complex with L7 and alone.**

| H 1 | H 2 | $\sigma_{IS}$ (Mix) | $\sigma_{IS}$ (Single) | Increase [%] | $\tau_c$ (Single) [$10^{-10}$ $s^{-1}$] | $\tau_c$ (Mix) [$10^{-10}$ $s^{-1}$] |
|---|---|---|---|---|---|---|
| Arg(8)-βH2 | Arg(8)-βH3 | 0.644 | 0.346 | 46 | 3.5 | 4.8 |
| Thr(2)-NH | Gln(1)-αH | 0.048 | 0.023 | 53 | | |
| Arg(3)-NH | Thr(2)-αH | 0.107 | 0.212 | -98 | | |
| Thr(4)-NH | Arg(3)-αH | 0.100 | 0.101 | -1 | | |
| Asn(5)-NH | Thr(4)-αH | 0.150 | 0.063 | 58 | | |
| His(7)-NH | Thr(6)-αH | 0.218 | 0.223 | -2 | | |
| Arg(8)-NH | His(7)-αH | 0.173 | 0.122 | 29 | | |
| Asp(9)-NH | Arg(8)-αH | 0.229 | 0.125 | 45 | | |
| Gly(10)-NH | Asp(9)-αH | 0.085 | 0.007 | 92 | | |
| His(7)-NH | Thr(6)-NH | 0.115 | 0.133 | -16 | | |
| Thr(6)-αH | Thr(6)-γH2 | 0.186 | 0.091 | 51 | | |

**Table 3.24. Rotational correlation times ($\tau_c$) and differences in cross-relaxation rate constants for protons in of L7, in complex with T20 and alone.**

| H 1 | H 2 | $\sigma_{IS}$ (Mix) | $\sigma_{IS}$ (Single) | Increase [%] | $\tau_c$ (Single) [$10^{-10}$ $s^{-1}$] | $\tau_c$ (Mix) [$10^{-10}$ $s^{-1}$] |
|---|---|---|---|---|---|---|
| Tyr(3)-βH2 | Tyr(3)-βH3 | 0.783 | 0.490 | 37 | 4.1 | 5.4 |
| Asp(2)-NH | Glu(1)-αH | 0.126 | 0.096 | 24 | | |
| Asp(2)-NH | Asp(2)-αH | 0.190 | 0.125 | 34 | | |
| Glu(6)-NH | Val(5)-αH | 0.206 | 0.126 | 39 | | |
| Tyr(3)-NH | Asp(2)-αH | 0.097 | 0.121 | -25 | | |
| Val(5)-NH | Glu(4)-αH | 0.245 | 0.082 | 67 | | |
| Gly(8)-NH | Glu(7)-αH | 0.074 | 0.036 | 52 | | |
| Glu(4)-NH | Tyr(3)-αH | 0.172 | 0.131 | 24 | | |

It is of course possible that the rotational correlation time was increased by a difference in the viscosity of the solvent. It is thus important to state that the samples were acquired at the same temperature and pH and that any small differences in the two are not likely to give rise to the rather big increase in intensity. The possibility cannot completely be disregarded though.

### Titration

To verify that a plausible interaction was present a titration study was performed, but no binding constant was obtainable, which would also be impossible to be studied with NMR due to a $k_D$ far below a minimum of $10^{-5}$ suggested from the literature.[109] It was observed that the change in relative chemical shifts of the Ligand 7 was dependent on the presence of Target 20, by performing a titration of L7 with T20, as seen in Figure 3.26, and afterwards diluting the sample to regain the original shifts.

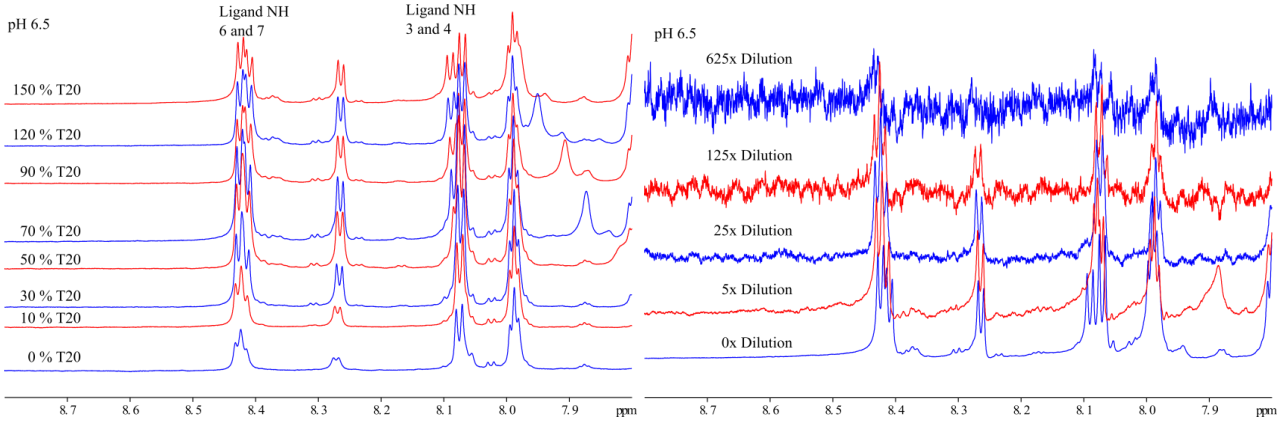



**Figure 3.26. Left: 1D $^1$H spectra of the titration of L7 with T20 at pH 6.5. Note that the resonances of T20 were not observed at this pH due to exchange with the solvent. Right: Dilutions of the sample after titration.**

### Structures from MD simulations

The structures of Target 20 and Ligand 7 were identified to be too flexible to fit conformers from unconstrained simulations to the data due to too many and too different conformers. It was clear that some degree of order was present from multiple NH-$C_\alpha$H distances corresponding or approaching those of $\beta$-strands (2.2 Å) and long distances for both compounds, given in Table 3.25, and constrained simulations were used in order to get backbone structures which correlated well to the NOE data.[110]

**Table 3.25. Experimental distances of $NH_i$-$C_\alpha H_{i-1}$ for T20 and L7, as individual peptides and when mixed. Determined by setting the distance of a diastereotopic proton pair to 1.78 Å (T20: Arg(8)-β, L7: Tyr(3)-β).**

| H1 | H2 | Dist. [Å] | H1 | H2 | Dist. [Å] |
|---|---|---|---|---|---|
| **Target 20, individual peptide** | | | **Target 20, mixture** | | |
| Thr(2)-NH | Gln(1)-αH | 3.21 | Thr(2)-NH | Gln(1)-αH | 2.94 |
| Arg(3)-NH | Thr(2)-αH | 2.42 | Arg(3)-NH | Thr(2)-αH | 2.50 |
| Arg(3)-NH | Thr(4)-NH | 3.29 | - | - | - |
| Thr(4)-NH | Arg(3)-αH | 2.38 | Thr(4)-NH | Arg(3)-αH | 2.54 |
| Asn(5)-NH | Thr(4)-αH | 2.43 | Asn(5)-NH | Thr(4)-αH | 2.37 |
| His(7)-NH | Thr(6)-αH | 2.30 | His(7)-NH | Thr(6)-αH | 2.22 |
| His(7)-NH | Thr(6)-NH | 2.55 | His(7)-NH | Thr(6)-NH | 2.38 |
| Arg(8)-NH | His(7)-αH | 2.16 | Arg(8)-NH | His(7)-αH | 2.31 |
| Asp(9)-NH | Arg(8)-αH | 2.46 | Asp(9)-NH | Arg(8)-αH | 2.21 |
| Gly(10)-NH | Asp(9)-αH | 3.19 | Gly(10)-NH | Asp(9)-αH | 2.60 |
| **Ligand 7, individual peptide** | | | **Ligand 7, mixture** | | |
| Asp(2)-NH | Glu(1)-αH | 2.44 | Asp(2)-NH | Glu(1)-αH | 2.44 |
| Tyr(3)-NH | Asp(2)-αH | 2.35 | Tyr(3)-NH | Asp(2)-αH | 2.55 |
| Glu(4)-NH | Tyr(3)-αH | 2.32 | Glu(4)-NH | Tyr(3)-αH | 2.31 |
| Glu(4)-NH | Val(5)-NH | 2.96 | Glu(4)-NH | Val(5)-NH | 2.98 |
| Val(5)-NH | Glu(4)-αH | 2.51 | Val(5)-NH | Glu(4)-αH | 2.18 |
| Glu(6)-NH | Val(5)-αH | 2.33 | Glu(6)-NH | Val(5)-αH | 2.25 |
| Glu(6)-NH | Val(5)-NH | 3.04 | Glu(6)-NH | Val(5)-NH | 2.89 |
| Gly(8)-NH | Glu(7)-αH | 2.88 | Gly(8)-NH | Glu(7)-αH | 2.66 |
| Gly(8)-NH | Glu(7)-NH | 2.99 | - | - | - |

The resulting backbone structures fit well to the data and are shown in Figure 3.27 but it is imperative to stress that these are representations of the conformational space.

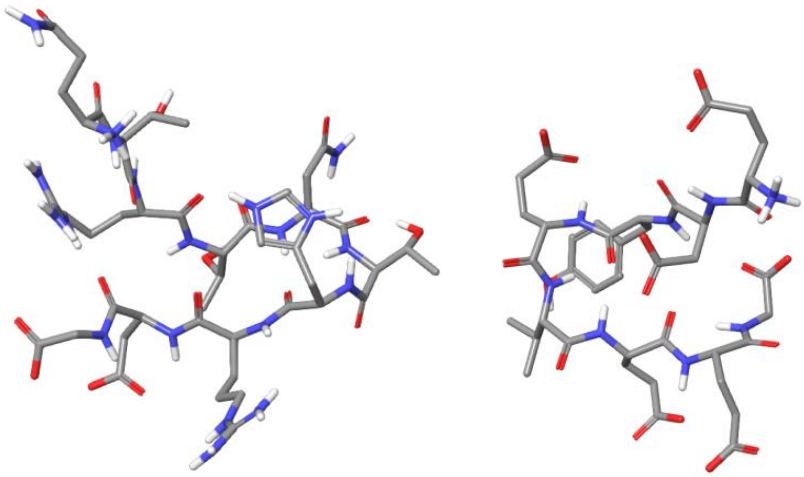


**Figure 3.27. Representative 3D structure of T20 (left) and L7 (right) from constrained simulations using NOE-derived distances.**

The two structures both exhibited a turn, in good agreement with NH-NH distances from the NOE data. The two structures were found undergo small adaptations upon interacting, but the changes were minor, and largest for L7, again in good correlation to the NMR data, illustrated in Figure 3.28. Four water molecules were entrapped in a cavity between the two structures in the simulations, which established a hydrogen bonding network with both T20 and L7, while electrostatic interactions and topological complementarity between the structures was observed along the rim of the cavity. Note that the water molecules are not shown in Figure 3.28. The

interaction was thus based on side-chain to side-chain interactions, and water mediated hydrogen bonds, which could be the reason for the lack of intermolecular NOEs. Another suggestion could be a high on-off rate which could hamper the buildup of NOE signal.

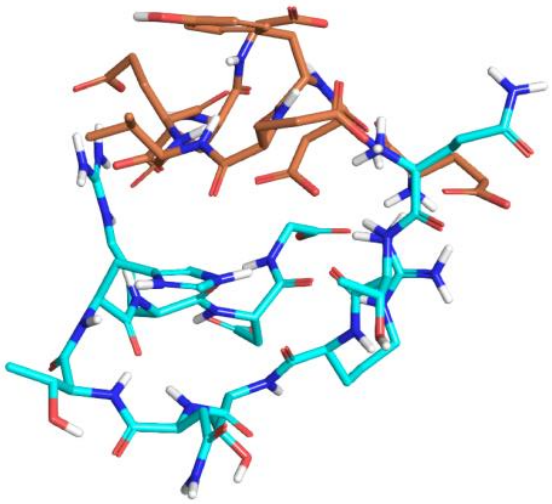


**Figure 3.28. The interaction between Target 20 (teal) and Ligand 7 (orange) from MD simulations using constraints from NOE distances. Interactions are dominated by charged side-chain interactions. Water molecules in the cavity of the peptides omitted.**

### Conclusion

Two peptides were investigated by NMR spectroscopy to obtain information of a highly specific binding interaction, suggested from binding data. It was not possible to identify the interaction from inter-peptide NOEs, but several indices suggested that the peptides interact, including shift in proton chemical shifts upon mixing, a shift in the rotational correlation time and small changes in NOE intensities. The peptides were modelled based on distances obtained from the relative NOE intensities, and two complementing structures were identified, which fitted the backbone data well, and also theoretically bind well, primarily by interactions of the outermost parts of side-chains. The fact that the peptides contained a potential for much more flexibility, led to the need for constrained simulations to obtain good structures, and the given structures are given as representatives of the conformational space the peptides occupy.

## 3.4 Other compounds

*This and the following Section cover a couple of projects and points from my external stay in the group of Craig Butts at the University of Bristol. This first section covers the usage of NOEs coupled to DFT optimized or clustered structures to gain knowledge of conformer populations. The NOE work is based on the PANIC approach, covered in Section 2.1.5.*

### 3.4.1 Population analysis of quinine

The alkaloid quinine has been thoroughly investigated due to interesting biological activities and usage in chiral synthesis.[111–114] The conformational populations have previously been solved from Gibbs free energies from Hatree-Fock and DFT calculation and the *J*-coupling constants of H8-H9.[115,116] The related compound cinchonidine, varied only by missing the methoxy group at the aromatic position 6’ (Figure 3.29), was recently investigated using DFT calculations and NOEs,

although a limited set of NOEs was used (only < 3 Å included) and the conclusions were negative in correlating the theoretical energy populations to the NOE populations.[117] It could thus be interesting to see if a conclusion could be reached for quinine by the usage of the PANIC approach and a full NOE dataset. When investigating the spectra only three distances above 3 Å were available, all of which were fixed distances due to structural rigidity, so the difference between datasets is not expected to be significant.

The group of Craig Butts had previously investigated quinine and determined the conformers from relative conformer energies at a B3LYP/6-31g(d) level of theory in implicit chloroform, and the conformers are compared to those reported by Urakawa *et al.* for cinchonidine, see Table 3.27 and Figure 3.30.[117]

**Table 3.26. Distances used in the conformational analysis of quinine. [a]Reference**

| H1 | H2 | Exp. Dist. [Å] |
|---|---|---|
| 2' | 3' | 2.42 |
| 9 | 3' | 2.85 |
| 9 | 5' | 2.11 |
| 9 | 6a | 2.51 |
| 9 | 8 | 2.55 |
| 9 | 7a | 2.82 |
| 6a | 6b | 1.79[a] |
| 6a | 5a | 2.31 |
| 6a | 5b | 3.10 |
| 8' | 7' | 2.54 |
| 3 | 2a | 2.36 |
| 3 | 6b | 3.14 |
| 3 | 4 | 2.55 |
| 3 | 5b | 2.49 |
| 8 | 3' | 2.76 |
| 8 | 5' | 2.36 |
| 2a | 2b | 1.71 |
| 8 | 7a | 2.82 |
| 8 | 7b | 2.29 |
| 2b | 3 | 2.92 |
| 6b | 5a | 3.11 |
| 6b | 5b | 2.38 |

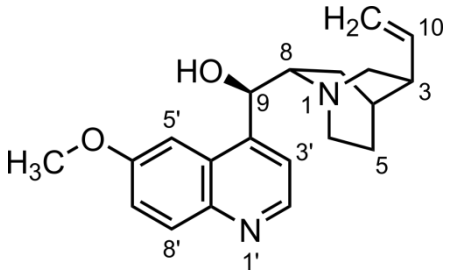



**Figure 3.29. Structure of quinine.**

**Table 3.27. Populations in % from the literature and from the group of Craig Butts.**

| | Dihedral Angles (°) | | | % Population | |
|---|---|---|---|---|---|
| Conformer | $\tau_1$ | $\tau_2$ | $\tau_3$ | Urakawa[117] | Butts[II] |
| C1 | 249 | 50 | 175 | 10 | 8 |
| C2 | 60 | 60 | 180 | 4 | 2 |
| C7 | 17 | 53 | 173 | 11 | 10 |
| O3 | 99 | 150 | 278 | 72 | 73 |
| O4 | 260 | 140 | 272 | 0 | 4 |
| O8 | 20 | 150 | 278 | 0 | 3 |
| O10 | 102 | 268 | 38 | 3 | 0 |

Most of the distances were between rigid parts of the molecule, and should not influence the determined populations. The non-rigid distances of the seven relevant conformers which influenced the obtained populations are given in Table 3.28. All distances are included in Appendix A4.

[II] Unpublished work of the group of Dr. Craig Butts

**Table 3.28. Experimental distances between non-rigid proton pairs and distances determined from the seven relevant conformers of quinine. Structures were optimized using MPW1PW91/6-311+g(d,p).**

| H1 | H2 | Exp. | C1 | C2 | O3 | C7 | O8 | O4 | O10 |
|---|---|---|---|---|---|---|---|---|---|
| 9 | 3' | 2.9 | 2.28 | 3.68 | 3.65 | 3.55 | 3.62 | 3.62 | 3.61 |
| 9 | 5' | 2.1 | 3.79 | 2.09 | 2.13 | 2.34 | 2.29 | 2.29 | 2.07 |
| 9 | 6a | 2.5 | 2.13 | 2.16 | 2.67 | 2.23 | 2.69 | 2.69 | 3.97 |
| 9 | 8 | 2.6 | 3.04 | 3.04 | 2.56 | 3.05 | 2.53 | 2.53 | 2.24 |
| 9 | 7a | 2.8 | 2.60 | 2.71 | 3.63 | 2.44 | 3.65 | 3.65 | 2.85 |
| 8 | 3' | 2.7 | 2.75 | 2.72 | 3.83 | 2.33 | 2.19 | 2.19 | 2.45 |
| 8 | 5' | 2.4 | 2.41 | 2.22 | 2.44 | 3.98 | 3.70 | 3.70 | 3.30 |

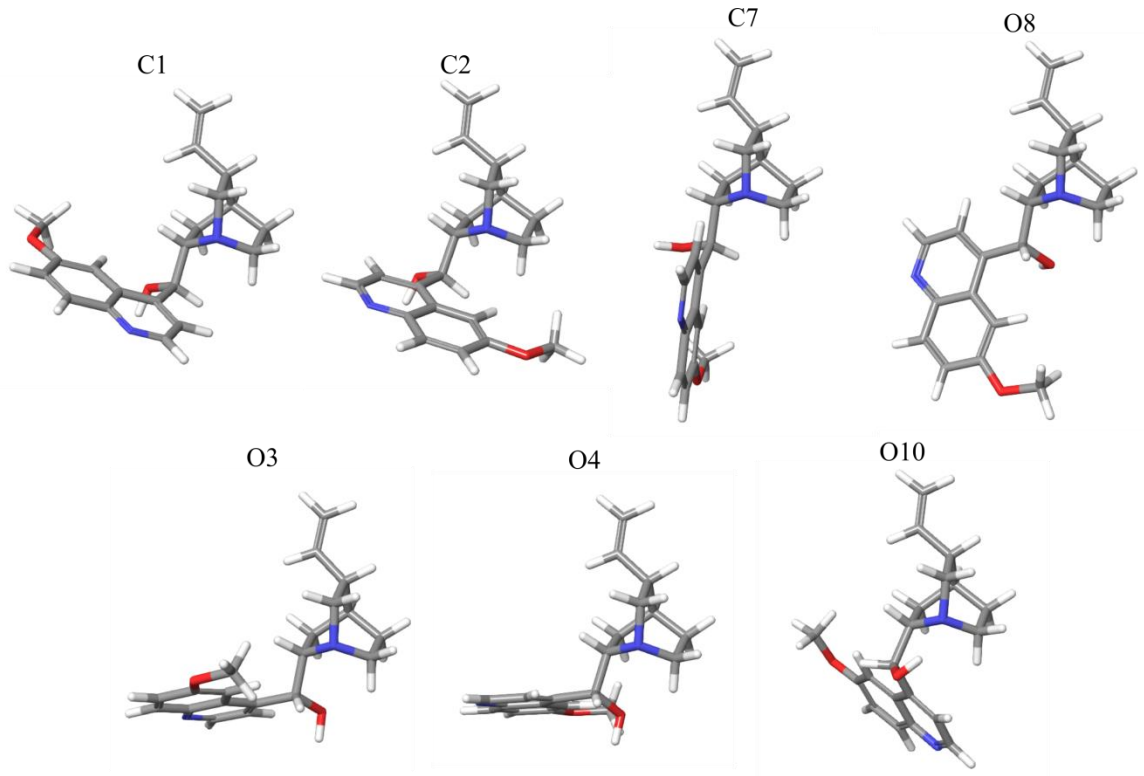



**Figure 3.30. The most abundant conformers of quinine (B3LYP/6-31g(d) and literature[117]).**

Initially the populations of the six conformations identified in the group of Craig Butts were compared to the NOE data. The populations of single conformers were fixed individually and the remaining conformers were optimized to give the best fit to the data. It was realized that the MAE was not discriminative, since the MAEs of most populations were generally low. Instead the maximum difference for the populations was used, as a better discriminator of the actual correlation between NOE distances and 3D structures see Figure 3.31.

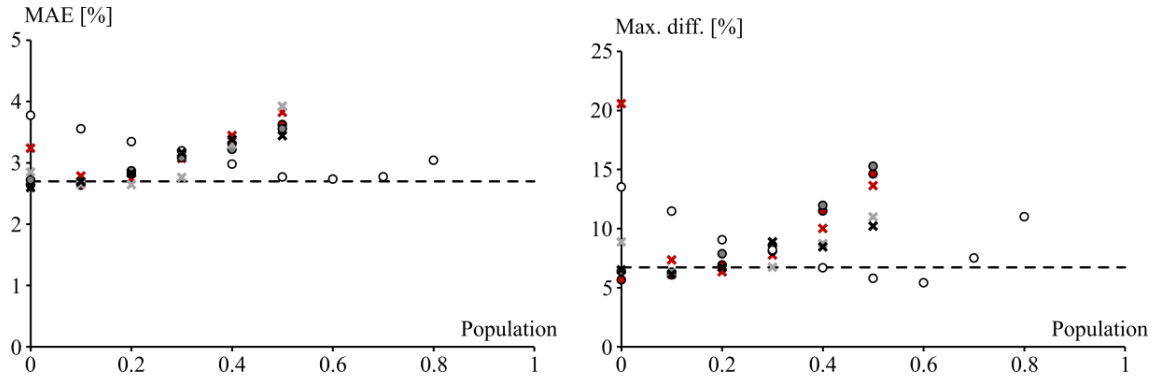



**Figure 3.31. Population analyses, by fixing a conformer population and optimizing the rest, of the six conformers C1(×), C2(×), C7(×), O3(○), O4(●) and O8(●). The dotted line is the value using only C1, C7 and O3.**

NOE distances dictated that the O3 conformer was indeed the dominating one, in accordance with all DFT calculations, and that the remaining conformers were populated 0-20 %. Due to overlap in the conformational space between the conformers only three conformers were needed in order to obtain a good fit to the data, to a good approximation, namely O3, C1 and C7. There was no solid experimental evidence that the rest of the conformations are present in solution, probably due to structural similarities within open and closed conformations.

When adding the seventh conformation, O10, to the available structures as given by the population analysis of Urakawa *et al.*, it led to a slightly better fit, considering the MAE, while the maximum difference in the distances was almost unaffected. The NOE populations were generally worse compared to the theoretical populations, and akin to C2, O4 and O8, the conformer was not proven needed in the fitting due to a low dependence on the inclusion of the conformer, Figure 3.32.

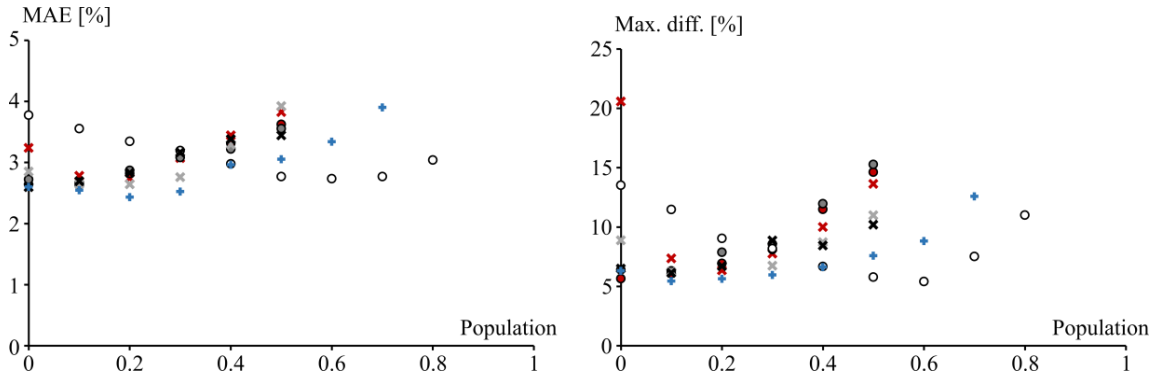



**Figure 3.32 Population analyses, by fixing a conformer population and optimizing the rest, of the seven conformers C1(×), C2(×), C7(×), O3(○), O4(●), O8(●) and O10(+).**

In Table 3.29 the conformational populations using the different sets of structures are shown, which may be compared to the data in Table 3.27.

**Table 3.29. Experimental population of the conformers from NOE distances, using different sets of conformers as indicated.**

| | Exp. Population [%] | | |
|---|---|---|---|
| Conformer | 6 structures | 7 structures | Essential 3 |
| C1 | 15 | 16 | 13 |
| C2 | 7 | 16 | - |
| C7 | 14 | 4 | 21 |
| O3 | 59 | 41 | 66 |
| O4 | 4 | 0 | - |
| O8 | 1 | 0 | - |
| O10 | - | 23 | - |
| MAE [%] | 2.54 | 2.43 | 2.70 |
| Max diff [%] | 6.31 | 5.45 | 6.73 |

In conclusion it was difficult to determine the populations from the data, as the introduction of the O10 conformer led to large differences between experimental and theoretical populations. The problems arise from overlaps of the conformational space between conformers, leading to few distances that may actually differentiate conformers. There were seven distances identified between rotatable parts of the structure, which, compared to up to seven conformers, leads to an underdetermined system when trying to determine a population of each conformer. It might be more interesting to fit the dihedral angles around the rotatable bonds, which is seen in Figure 3.33.

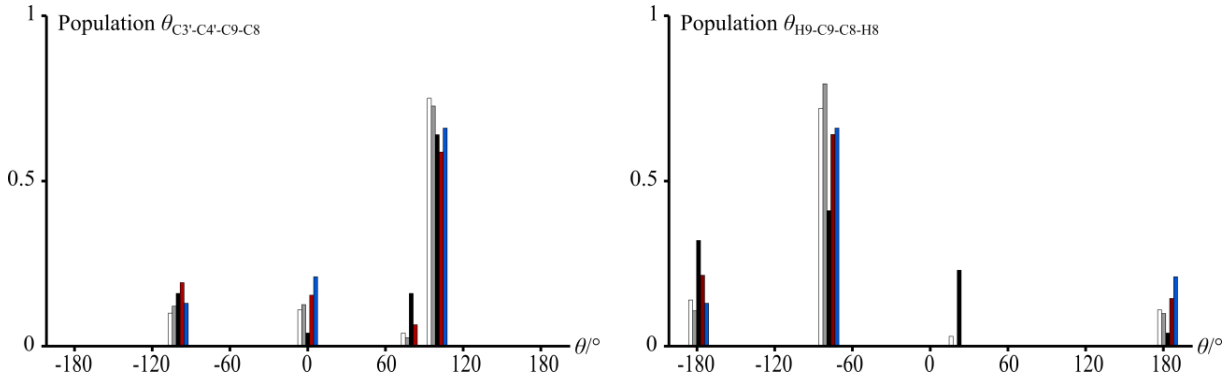



**Figure 3.33. The population of dihedral angles associated with the rotatable bonds in quinine. White: Urakawa (energy), grey: Butts (energy), black: 7 structures (NOE), red: 6 structures (NOE), blue: 3 structures (NOE).**

The only really different populations were observed when including the seventh conformer which seemed to be severely overestimated. The populations of quinine are revisited in Chapter 0.

## 3.5 Inclusion of correlation time in NOE calculations

*On my external stay I was involved in a project where the goal was to utilize calculation of correlation times to improve on the accuracy of NOE calculations for small organic molecules. This was tried achieved in multiple ways, I will include only the parts where I was directly involved.*

### 3.5.1 Theory

A major assumption in the ISPA methodology, is equality of correlation time ($\tau_c$) for all nuclei pairs, it would seem that including relative correlation times would increase the accuracy of the calculations, as proposed in the literature.[15,118–122]

The actual parameter of interest is the effective rotational correlation time, which covers the overall rotational correlation time of the molecule ($\tau_o$) and the relative intramolecular movements ($\tau_s$) as given in (3.7).[118]

$$\frac{1}{\tau_c} = \frac{1}{\tau_o} + \frac{1}{\tau_s} \tag{3.7}$$

Note that the effective correlation time is also called $\tau_e$ or $\tau_i$ in the literature but covers the same parameter.[118,119]

Values of $\tau_c$ may be obtained in four different ways:[7,123]

1. By estimation from the size of the molecule (correlate with molecular weight of proteins).
2. By estimating correlation times from $T_1$ relaxation (only possible for $^{13}C$).
3. Utilizing NOE and ROE spectra at the same field strength.
4. Comparing NOEs at two field strengths.

The first two methods will not be discussed while the formula needed for the two latter methods are included in equations (3.8) to (3.16), starting from the differences in NOE and ROE observables in (3.8) and (3.9).[7,118–120] Note a small overlap with theory in Chapter 2.

$$\sigma_{NOE} = K \cdot r^{-6} \cdot [6J(2\omega) - J(0)] = k \cdot r^{-6} \cdot \tau_c \left( \frac{6}{1 + 4\omega^2 {\tau_c}^2} - 1 \right) \tag{3.8}$$

$$\sigma_{ROE} = K \cdot r^{-6} \cdot [3J(\omega) + 2J(0)] = k \cdot r^{-6} \cdot \tau_c \left( \frac{3}{1 + \omega^2 {\tau_c}^2} + 2 \right)$$
$$K = \left( \frac{\mu_0}{4\pi} \right)^2 \frac{\hbar^2 {\gamma_H}^4}{10} \tag{3.9}$$

In practice the most utilized ROE experiment is the T-ROE, due to the less demanding spin lock utilized. In T-ROE the magnetization may be viewed as shared between longitudinal and transverse during the mixing time as given in (3.10).[14,124]

$$\sigma_{T-ROE} = \frac{\cos\theta_i \cos\theta_j \, \sigma_{NOE} + \left(1 + \sin\theta_i \sin\theta_j\right)\sigma_{ROE}}{2} \tag{3.10}$$

With $\tan\theta = (\omega_i - \omega_T)/\gamma B_1$. $\omega_i$ is the chemical shift of $i$ in rad/s (or Hz, need to match $\gamma B_{1)}$, $\omega_T$ is the chemical shift of spin-lock transmitter offset in rad/s (or Hz), $\gamma B_1$ is the spin-lock field in rad/s (or Hz). If the assumption is made that $\theta_i$ and $\theta_j \rightarrow 0$, then $\cos\theta_i$ and $\cos\theta_j \rightarrow 1$ and $\sin\theta_i$ and $\sin\theta_j \rightarrow 0$. This is the most abundant approach used in the literature on the subject.[15,118–122] This leads to (3.11).

$$\begin{aligned} \sigma_{T-ROE} &= \frac{K}{2} \cdot r^{-6} \cdot [6J(2\omega) + 3J(\omega) + J(0)] \\ &= \frac{K}{2} \cdot r^{-6} \cdot \tau_c \left( \frac{6}{1 + 4\omega^2 {\tau_c}^2} + \frac{3}{1 + \omega^2 {\tau_c}^2} + 1 \right) \end{aligned} \tag{3.11}$$

It is fairly obvious that this assumption is not valid for most systems. If one only wants to utilize the T-ROE intensities for distance calculations the approximation may work just fine due to the aforementioned $r^{-6}$ relationship, but the influence of $\tau_c$ may be problematic. With a $\gamma B_1$ of 5500 Hz at 500 MHz, a proton 4 ppm from the offset with its "partner" proton on the offset resonance will

need a correction of $\sigma_{ROE}$ of ≈ 5 %. If the two protons reside at ± 4 ppm of the offset resonance a correction of $\sigma_{ROE}$ of ≈ 25 % is needed.

From equations (3.8) to (3.11) the ratios of NOE to (T-)ROE may be given as equation (3.12) and (3.13).[119,125]

$$\frac{\sigma_{NOE}}{\sigma_{ROE}} = \frac{5 + \omega^2\tau_c{}^2 - 4\omega^4\tau_c{}^4}{5 + 22\omega^2\tau_c{}^2 + 8\omega^4\tau_c{}^4} \quad \textbf{(3.12)}$$

$$\frac{\sigma_{NOE}}{\sigma_{T-ROE}} = \frac{10 + 2\omega^2\tau_c{}^2 - 8\omega^4\tau_c{}^4}{10 + 23\omega^2\tau_c{}^2 + 4\omega^4\tau_c{}^4} \quad \textbf{(3.13)}$$

Note again that the latter is only true if $\theta_i$ and $\theta_j \rightarrow 0$. Still, even with this restriction, this is what is utilized in the literature.[118,119,125] Experimental data that suggests the need for the correction is presented in Appendix A9 in the form of different, albeit correctable, obtained T-ROE intensities at different pulsed field strengths and offsets. The relative NOE to ROE intensities are illustrated in Figure 3.34.

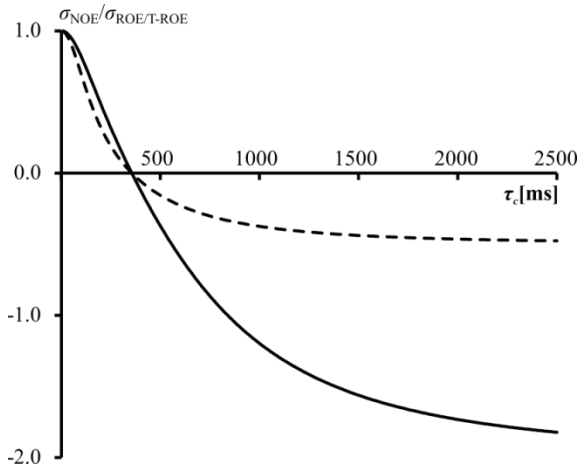



**Figure 3.34. The theoretical relative cross-relaxation rate ($\sigma$) of NOE and ROE (—) or T-ROE (- -) experiments at 500 MHz for a proton pair. Calculated using (3.12) and (3.13).**

If NOE and T-ROE data is acquired the corrected ROE part of T-ROE can be extracted by (3.14)

$$\sigma_{ROE} = \frac{2\sigma_{T-ROE} - \cos\theta_i \cos\theta_j \, \sigma_{NOE}}{\left(1 + \sin\theta_i \sin\theta_j\right)} \quad \textbf{(3.14)}$$

From (3.12) and (3.13) the correlation time may be calculated by (3.14) and (3.15) for $\sigma_{NOE} > 0$ and defining that $\sigma_{NOE}/\sigma_{(T\text{-})ROE} = \nabla\sigma$.

$$\tau_c{}^{NOE/ROE} = \frac{1}{4}\frac{\sqrt{2}\sqrt{(2\nabla\sigma + 1)(-22\nabla\sigma + 1 + 3\sqrt{36\nabla\sigma^2 + 4\nabla\sigma + 9}}}{(2\nabla\sigma + 1)\omega} \quad \textbf{(3.15)}$$

$$\tau_c{}^{NOE/T-ROE} = \frac{1}{4}\frac{\sqrt{2}\sqrt{(\nabla\sigma + 2)(-23\nabla\sigma + 2 + 3\sqrt{41\nabla\sigma^2 - 28\nabla\sigma + 36}}}{(\nabla\sigma + 2)\omega} \quad \textbf{(3.16)}$$

For two field NOE analysis only (3.8) is needed. If data is acquired at field *x* and *y*, and assuming $\sigma_{NOE,x} > \sigma_{NOE,y}$, equations (3.17) to (3.19) may be used.[122]

$$\sigma_{NOE,x} = K \cdot r^{-6} \cdot \tau_c \left( \frac{6}{1 + 4{\omega_x}^2 {\tau_c}^2} - 1 \right) \quad \textbf{(3.17)}$$

$$\sigma_{NOE,y} = K \cdot r^{-6} \cdot \tau_c \left( \frac{6}{1 + 4{\omega_y}^2 {\tau_c}^2} - 1 \right) \quad \textbf{(3.18)}$$

$$\frac{\sigma_{NOE,x}}{\sigma_{NOE,y}} = \frac{(4{\omega_x}^2 {\tau_c}^2 - 5)(4{\omega_y}^2 {\tau_c}^2 + 1)}{(4{\omega_x}^2 {\tau_c}^2 + 1)(4{\omega_y}^2 {\tau_c}^2 - 5)} \quad \textbf{(3.19)}$$

From this the correlation time may be determined.

After the correlation time is determined the distances need to be calculated. There are different possibilities in doing this; first by using (3.8) to (3.11) directly as all parameters but the distance is known. Another method is using a PANIC like approach, exemplified in (3.20).

$$r_i = r_{ref} \left( \frac{\sigma_i}{\sigma_{ref}} \frac{\tau_{c,ref} \left( \frac{6}{1 + 4\omega^2 {\tau_{c,ref}}^2} - 1 \right)}{\tau_{c,i} \left( \frac{6}{1 + 4\omega^2 {\tau_{c,i}}^2} - 1 \right)} \right)^{-1/6} \quad \textbf{(3.20)}$$

It should be noted that in most cases where correlation times are used, they are not included in the distance calculations but compared and deemed identical. It is thus used more as a validation that the ISPA approach is appropriate.[119,125]

For the sake of inclusion, the last major equation which should be mentioned in regard to rotational correlation time is based in the rotational correlation times for 13C which may be calculated from the $T_1$ relaxation time as shown in (3.21).[7]

$$\frac{1}{T_1(^{13}\mathrm{C})} = \left(\frac{\mu_0}{4\pi}\right)^2 N\hbar^2 {\gamma_H}^2 {\gamma_C}^2 r^{-6} \tau_c \quad \textbf{(3.21)}$$

Where *r* is the C-H distance (~1.1 Å) and *N* is the number of attached protons. The usage of this formula will of course approximate that the, at least relative, rotational correlation times are equal for bound protons and carbons, which may be a good approximation, though it was not thoroughly investigated.

### 3.5.2 Inclusion of correlation times for small molecular NMR

To test if the inclusion of the rotational correlation time would improve the correlation between experimental NOE derived distances and distances from 3D structures of small organic molecules in organic solvents, three compounds were used; rigid, and more importantly here, spherical strychnine, rigid but flat progesterone and more flexible quinine. While data of strychnine was only acquired in chloroform, data for the remaining two compounds were also acquired in DMSO.

**Strychnine**

Due to the spherical structure and excellent fit between experimental and 3D structure distances of strychnine it was not anticipated that including the rotational correlation time would lead to an overall better structural distance correlation. Experimentally, many of the observed T-ROEs were of lower intensity than the NOE counterparts, which is not theoretically possible, if assuming that the formulae in the previous section are valid. This is problematic and results in a significantly smaller

number of NOEs which are compared in Table 3.30. The fit between experimental and 3D structure distances was not improved but seem to be corrected more or less at random.

**Table 3.30. Theoretical and experimental distances between protons in strychnine using NOEs or $\tau_c$ corrected distances using the NOE/T-ROE methodology ($\tau_m$ = 300 ms). Structure optimized to MPW1PW91/6-311+g(d,p). Only distances where NOE<T-ROE intensities are included. [a]Reference distance.**

| H1 | H2 | 3D struct. [Å] | Experimental distances [Å] | | | | $\tau_c$ [ps] |
|---|---|---|---|---|---|---|---|
| | | | NOE | Diff [%] | NOE/T-ROE | Diff [%] | |
| 4 | 3 | 2.49 | 2.55 | 2.26 | 2.52 | 1.10 | 31.1 |
| 13 | 8 | 3.00 | 3.17 | 5.74 | 3.27 | 9.09 | 41.5 |
| 13 | 12 | 2.35 | 2.34 | 0.24 | 2.45 | 4.27 | 45.6 |
| 13 | 14 | 2.41 | 2.38 | 1.32 | 2.55 | 5.68 | 55.1 |
| 13 | 15a | 2.24 | 2.31 | 2.73 | 2.28 | 1.65 | 31.3 |
| 15a | 13 | 2.24 | 2.23 | 0.71 | 2.29 | 1.83 | 39.8 |
| 15a | 15b | 1.75[a] | 1.75 | 0.20 | 1.75 | 0.20 | 33.5 |
| 15a | 14 | 2.54 | 2.52 | 0.89 | 2.55 | 0.09 | 35.8 |
| 15a | 16 | 2.47 | 2.44 | 0.98 | 2.48 | 0.66 | 37.5 |
| 18b | 8 | 2.30 | 2.19 | 4.79 | 2.22 | 3.43 | 36.9 |
| 18b | 18a | 1.77 | 1.82 | 3.01 | 1.79 | 1.34 | 30.1 |
| 18b | 22 | 3.12 | 2.99 | 4.30 | 3.20 | 2.35 | 54.5 |
| | | | MAE[%] | 2.26 | | 2.64 | |

It was realized that the problem was the low values of the rotational correlation time for strychnine. Thus the theoretical differences between NOE and T-ROE intensities were lower than the experimental error. This is in good agreement with the observations; the values of $\tau_c$ were probably in the correct interval, but small differences were not translated well to distances. This problem is further addressed in explained in Table 3.31. Remember that a resulting distance error of 2 or 5 % translates to the intensity theoretically being either ~11 or 25 % low or high, or ~13 or 36 % too high. The values or 2 and 5 % were chosen as the average error approaches 2 % for strychnine and the maximum single distance error approaches 5 %. The errors in observed intensities were thus easily above the differences from differences in rotational correlation times. For comparison an intensity error of 2 % translates to a difference in 0.33 % in distances, which is the precision needed for NOE calculations to obtain a better fit in the case of strychnine.

**Table 3.31. Theoretical PANIC intensities of different distances at 500 MHz assuming $\tau_c$ = 30 ps.**

| r | $\eta_{NOE}$ | $\eta_{T\text{-}ROE}$ | Diff [%] |
|---|---|---|---|
| 1.75 | 128.8 | 131.2 | 1.9 |
| 2.25 | 30.2 | 30.8 | 2.0 |
| 2.50 | 13.6 | 13.9 | 2.2 |

Slightly discouraged by this result, the fact that strychnine is spherical and that chloroform has a low viscosity and thus expected short $\tau_c$, prompted for the investigation into other compounds which could be investigated in multiple solvents, chosen to be chloroform and DMSO due to large differences in viscosity. Salts of strychnine may be dissolved in DMSO, which was attempted, but low solubility gave inconclusive results that are not included.

**Progesterone**

The second compound to investigate was the steroid progesterone. The structure of progesterone is flat and asymmetrically extended in virtue of being a steroid. It was thus proposed that the rotational correlation time would differ significantly dependent on whether the distance vector would lie along the long (L or 1), short (S or 2) axis or in plane (3), see Figure 3.35 for clarification.

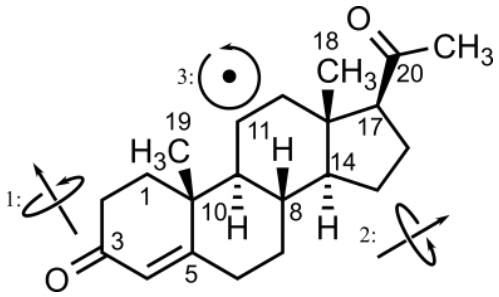



**Figure 3.35. The structure of progesterone, with the three proposed major rotation axis are indicated. In the following the term "L" will cover 1 and 3, and "S" 2.**

A theoretical dependence was identified as two distinct groups, correlating to rotation 1 and 2, the first also including 3, with grouped distance errors. In practice it was troublesome to couple this to the actual rotational correlation time in chloroform, and only minor improvements were obtained when tested in DMSO.

**Chloroform**

The data obtained from chloroform is seen in Table 3.32. The direction of the distance vector were divided into long (L) and short (S) from assessment of a 3D structure (see Figure 3.35, where L = 1 and 3, and S = 2). The theoretically optimized experimental distances (Exp Theo) gives experimental distances from theoretically optimized values of $\tau_c$ of these two groups which identified to be L = 37 and S = 20 ps (from optimizations using the Excel solver function), which gives a correction factor of 1.10 between the two. The experimental distances were corrected by the $\tau_c$ determined from the NOE/ROE intensities (Exp. Corr.). The differences between experimental and theoretical distances for corrected compared to uncorrected data is given by the subtracting the error of the corrected data from the uncorrected (Impr.); positive is an improved and negative a decreased fit between experimental and 3D structure distances by correcting the data.

**Table 3.32. Progesterone in $CDCl_3$. NOESY and T-ROESY data used for calculation of $\tau_c$ ($\tau_m$ = 300 ms). Structure optimized to MPW1PW91/6-311+g(d,p). *NOE/ROE was not solvable as $I_{NOE} > I_{ROE}$. H7b-H7a was used as reference. See text for further explanations.**

| H1 | H2 | Dir | Calc. | Exp. | Diff | Exp. Theo. | $\tau_c$ [ps] | Exp. Corr. | Impr. |
|---|---|---|---|---|---|---|---|---|---|
| 4 | 6a | L | 2.36 | 2.15 | 0.21 | 2.36 | 36 | 2.22 | 0.07 |
| 17 | 16a | S | 2.27 | 2.32 | -0.05 | 2.32 | 21 | 2.21 | -0.02 |
| 17 | 12a | L | 2.43 | 2.24 | 0.19 | 2.46 | 42 | 2.37 | 0.13 |
| 17 | 14 | L | 2.46 | 2.16 | 0.29 | 2.38 | 40 | 2.27 | 0.10 |
| 6a | 4 | L | 2.36 | 2.04 | 0.32 | 2.24 | 38 | 2.12 | 0.08 |
| 6a | 7b | L | 2.46 | 2.16 | 0.30 | 2.37 | 54 | 2.35 | 0.19 |
| 6a | 7a | L | 2.45 | 2.32 | 0.12 | 2.55 | 46 | 2.48 | 0.09 |
| 16b | 16a | S | 1.76 | 1.75 | 0.01 | 1.75 | 60 | 1.93 | -0.17 |
| 16b | 15b | L | 2.31 | 2.19 | 0.12 | 2.40 | 97 | 2.52 | -0.10 |
| 7b | 6b | L | 2.44 | 2.33 | 0.11 | 2.56 | 45 | 2.48 | 0.06 |
| 7b | 6a | L | 2.46 | 2.20 | 0.26 | 2.41 | 63 | 2.44 | 0.24 |
| 7b | 8 | S | 2.46 | 2.48 | -0.02 | 2.48 | 76 | 2.81 | -0.33 |
| 7b | 15b | L | 2.78 | 2.47 | 0.31 | 2.71 | 64 | 2.75 | 0.28 |
| 7b | 7a | S | 1.76 | 1.74 | 0.03 | 1.74 | 64 | 1.93 | -0.15 |
| 8 | 6b | L | 2.58 | 2.34 | 0.24 | 2.57 | 37 | 2.43 | 0.09 |
| 8 | 7b | S | 2.46 | 2.56 | -0.10 | 2.56 | 86 | 2.93 | -0.37 |
| 8 | 15b | L | 2.68 | 2.56 | 0.12 | 2.81 | -* | | |
| 15b | 16b | L | 2.31 | 2.16 | 0.15 | 2.37 | 76 | 2.45 | 0.01 |
| 15b | 7b | L | 2.78 | 2.34 | 0.44 | 2.57 | -* | | |
| 15b | 15a | S | 1.77 | 1.77 | 0.00 | 1.77 | -* | | |
| 15b | 8 | L | 2.68 | 2.48 | 0.20 | 2.72 | 61 | 2.75 | 0.13 |
| 7a | 6a | S | 2.45 | 2.43 | 0.02 | 2.43 | 27 | 2.41 | -0.03 |
| 7a | 7b | S | 1.76 | 1.76 | 0.00 | 1.76 | 29 | 1.76 | 0.00 |
| 7a | 15a | L | 3.05 | 2.64 | 0.41 | 2.90 | 92 | 3.03 | 0.39 |
| 9 | 1b | L | 2.29 | 2.08 | 0.20 | 2.29 | 42 | 2.20 | 0.11 |
| | | MAE [%] | | 6.63 | | 2.52 | | 5.50 | |
| | | St. dev. [%] | | 4.71 | | 2.03 | | 4.76 | |

The directionality of the distance vectors were assessed from a model and some may correspond more to a situation represented by rotation 3 in Figure 3.35. Still, when comparing the experimental to the theoretical distances, there seem to be a correlation between the assigned rotation axis and the fit. This was translated into theoretical $\tau_c$ values, and led to a much improved fit of the experimental distances. In practice the determined $\tau_c$ seem to be distributed at random, and though the experimental distances generally correlate better to the 3D structure, multiple distances had increased differences to the 3D structure. Also, as for strychnine, some $\tau_c$ values where not possible to identify due to a lower ROESY intensity compared to the relevant NOESY signal, which result in unsolvable equations. This may, as for strychnine, be caused by to the experimental error being larger than the theoretical difference between the NOESY and ROESY intensities, see Section 0 for further discussion.

**DMSO**

The data obtained from DMSO is found in Table 3.33.

**Table 3.33. Progesterone in DMSO-$d_6$. NOESY and T-ROESY data used for calculation of $\tau_c$ ($\tau_m$ = 300 ms). Structure optimized to MPW1PW91/6-311+g(d,p). H7b-H7a was used as reference. See text regarding $CDCl_3$ for further explanations.**

| H1 | H2 | Dir | Calc. | Exp. | Diff | Exp. Theo. | $\tau_c$ [ps] | Exp. Corr. | Impr. |
|---|---|---|---|---|---|---|---|---|---|
| 4 | 6a | L | 2.29 | 2.25 | 0.04 | 2.31 | 85 | 2.22 | -0.02 |
| 17 | 12b | S | 2.73 | 2.84 | -0.11 | 2.84 | 83 | 2.81 | 0.03 |
| 17 | 16a | S | 2.36 | 2.36 | 0.00 | 2.36 | 95 | 2.36 | 0.00 |
| 17 | 12a | L | 2.37 | 2.34 | 0.03 | 2.40 | 105 | 2.34 | 0.01 |
| 17 | 14 | L | 2.36 | 2.27 | 0.09 | 2.34 | 138 | 2.28 | 0.01 |
| 2b | 1a | S | 2.49 | 2.49 | 0.00 | 2.49 | 135 | 2.50 | -0.01 |
| 6b | 7b | L | 2.44 | 2.39 | 0.06 | 2.45 | 96 | 2.38 | 0.00 |
| 6b | 8 | L | 2.59 | 2.47 | 0.12 | 2.54 | 123 | 2.48 | 0.01 |
| 6b | 7a | S | 3.05 | 3.07 | -0.02 | 3.07 | 135 | 3.08 | -0.01 |
| 6a | 7b | L | 2.51 | 2.42 | 0.09 | 2.49 | 143 | 2.43 | 0.00 |
| 6a | 7a | L | 2.48 | 2.37 | 0.11 | 2.44 | 102 | 2.37 | 0.00 |
| 2a | 1a | S | 2.51 | 2.60 | -0.09 | 2.60 | 158 | 2.59 | 0.01 |
| 2a | 1b | L | 2.47 | 2.42 | 0.06 | 2.48 | 125 | 2.43 | 0.01 |
| 7b | 6b | L | 2.44 | 2.36 | 0.09 | 2.42 | 97 | 2.36 | 0.00 |
| 7b | 6a | L | 2.51 | 2.42 | 0.09 | 2.49 | 134 | 2.43 | 0.01 |
| 7b | 15b | L | 2.50 | 2.53 | -0.03 | 2.60 | 153 | 2.52 | 0.01 |
| 7b | 8 | S | 2.46 | 2.44 | 0.02 | 2.44 | 154 | 2.52 | 0.01 |
| 7b | 14 | S | 2.61 | 2.54 | 0.08 | 2.54 | 131 | 2.54 | 0.01 |
| 7b | 7a | S | 1.76 | 1.76 | 0.00 | 1.76 | 98 | 1.76 | 0.00 |
| | | MAE [%] | | 2.34 | | 1.40 | | 2.28 | |
| | | St. dev. [%] | | 1.36 | | 1.25 | | 1.32 | |

As for chloroform the directionality of the distance vectors were assessed from a model. In DMSO it is less obvious that one may correlate experimental distances to markedly different $\tau_c$. It was done though and led to an improvement in the already excellent fit between experimental distances and distances obtained from a 3D structure. In practice the determined $\tau_c$ seem to be distributed at random again, and the MAE is essentially unchanged. It is obvious that the higher viscosity of DMSO leads to a longer $\tau_c$ and that the experimental data fit better to the calculated data. This may be contributed to the signal intensity being less dependent of $\tau_c$ for this rotational time regime.

The same analysis may be done by the two-field (2F) methods, to much the same conclusion. A comparison of the $\tau_c$ values from NOE/ROE and 2F analysis is found in Table 3.34, and the rotational correlation times are markedly different, which is problematic. In most cases the rotational correlation times are in the same general regime though, and the differences between corrected distances are negligible. This is currently under assessment in the group of Dr. Craig Butts and will not be further elaborated on.

**Table 3.34. Progesterone in DMSO-$d_6$. Comparison of experimental $\tau_c$ from NOE/ROE and two-field analysis ($\tau_m$ = 300 ms). NOESY and T-ROESY data used for calculation of $\tau_c$ at 500 or 600 MHz using a spinlock field of 5452 Hz with an offset at 5 ppm.**

| H1 | H2 | $\tau_c$ [ps] | |
|---|---|---|---|
| | | NOE/ROE | 2F NOE |
| 4 | 6a | 85 | 88 |
| 17 | 16a | 95 | 63 |
| 17 | 12a | 105 | 97 |
| 17 | 14 | 138 | 130 |
| 6a | 7b | 143 | 90 |
| 6a | 7a | 102 | 165 |
| 2a | 1a | 158 | 124 |
| 2a | 1b | 125 | 119 |
| 7b | 6b | 97 | 255 |
| 7b | 6a | 134 | 165 |
| 7b | 8 | 154 | 134 |
| 7b | 7a | 98 | 93 |

## Quinine

From a general assessment of the structure of quinine, it was postulated that the two rigid systems might experience different $\tau_c$ and that the inter-rigid NOEs would most certainly experience a $\tau_c$ different than that of the rigid NOEs, see Figure 3.36 for clarification and the structure of quinine. Quinine was investigated in chloroform and DMSO, but in this text only the chloroform data is analyzed, as the outcome and conclusions are very similar for data acquired in the two solvents.

First the rigid parts of quinine was investigated, Table 3.35, and there seem to be a slight difference in the rotational correlation time.

**Table 3.35. Quinine in $CDCl_3$. NOESY and T-ROESY data used for calculation of $\tau_c$. Only rigid distances shown. Exp. Corr. is the distances corrected by the experimental $\tau_c$ values. Impr is the differences between experimental and theoretical distances for corrected minus uncorrected data; positive is an improved and negative a decreased fit between experimental and 3D structure distances by correcting the data. H6a-H6b was used as reference. Structures optimized to MPW1PW91/6-311+g(d,p).**

| H1 | H2 | $\tau_c$ [ps] | Exp. | Exp. Corr. | Calc. | Impr. |
|---|---|---|---|---|---|---|
| 2' | 3' | 188 | 2.48 | 2.45 | 2.49 | -0.03 |
| 8' | 7' | 178 | 2.56 | 2.54 | 2.52 | 0.02 |
| 3' | 2' | 186 | 2.43 | 2.40 | 2.49 | -0.03 |
| 3 | 2a | 166 | 2.39 | 2.39 | 2.33 | 0.00 |
| 3 | 6b | 162 | 3.13 | 3.14 | 3.32 | 0.01 |
| 3 | 4 | 149 | 2.53 | 2.55 | 2.49 | -0.03 |
| 3 | 5b | 156 | 2.48 | 2.49 | 2.44 | -0.02 |
| 2a | 2b | 167 | 1.72 | 1.72 | 1.77 | 0.00 |
| 2a | 3 | 163 | 2.32 | 2.33 | 2.33 | 0.01 |
| 8 | 7b | 137 | 2.26 | 2.30 | 2.31 | 0.03 |
| 6b | 6a | 171 | 1.79 | 1.79 | 1.76 | 0.00 |
| 2b | 2a | 174 | 1.69 | 1.69 | 1.76 | -0.01 |
| 2b | 3 | 207 | 3.02 | 2.92 | 2.93 | 0.07 |
| 6b | 5b | 148 | 2.36 | 2.38 | 2.36 | -0.02 |
| 6a | 6b | 169 | 1.78 | 1.78 | 1.76 | 0.00 |
| 6a | 5a | 143 | 2.29 | 2.32 | 2.35 | 0.03 |
| | MAE [%] | | 2.13 | 2.06 | | |
| | St. dev [%] | | 1.37 | 1.42 | | |

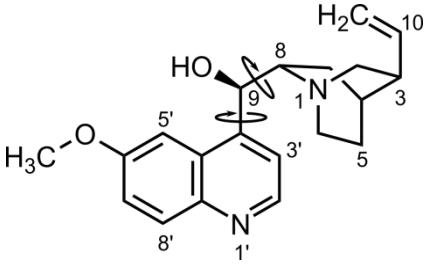



**Figure 3.36. The structure of quinine indicating the rotatable bonds connecting the rigid ring systems.**

The fit, in the form of the MAE, is slightly improved by incorporating the $\tau_c$ term in the calculations, but again the distances are not generally improved, and the structural fit is already excellent.

The inter-rigid distances and their $\tau_c$ are found in Table 3.36. It should be noted that the populations were optimized in both instances from the structures discussed in Section 3.4.1, and that the populations only varied slightly.

**Table 3.36. Quinine in CDCl3. NOESY and T-ROESY data used for calculation of τc ($\tau_m$ = 300 ms). Only non-rigid distances shown and conformer populations were optimized before comparison. Exp. Corr. is the distances corrected by the experimental $\tau_c$ values. Impr is the differences between experimental and theoretical distances for corrected minus uncorrected data; positive is an improved and negative a decreased fit between experimental and 3D structure distances by correcting the data. H6a-H6b was used as reference.**

| H1 | H2 | $\tau_c$ [ps] | Exp. | Exp. Corr. | Calc. | Impr. |
|---|---|---|---|---|---|---|
| 8 | 3' | 136 | 2.73 | 2.73 | 2.81 | 0.00 |
| 8 | 5' | 178 | 2.39 | 2.39 | 2.51 | 0.00 |
| 8 | 9 | 180 | 2.55 | 2.55 | 2.65 | 0.00 |
| 9 | 3' | 181 | 2.89 | 2.86 | 2.96 | -0.03 |
| 9 | 5' | 157 | 2.10 | 2.12 | 2.22 | 0.02 |
| 9 | 6a | 145 | 2.53 | 2.56 | 2.43 | -0.03 |
| 9 | 8 | 176 | 2.59 | 2.58 | 2.65 | -0.01 |
| 9 | 7a | 122 | 2.78 | 2.83 | 2.89 | 0.05 |
| 6a | 9 | 143 | 2.45 | 2.48 | 2.43 | -0.03 |
| | | MAE [%] | 3.35 | 3.49 | | |
| | | St. dev. [%] | 1.33 | 1.13 | | |

The $\tau_c$ was not markedly different for these flexible proton pairs than $\tau_c$ determined for the rigid parts, in contrast to our assumptions, and the MAE was actually slightly worsened by including the rotational correlation time in the calculations. When determining the populations, the difference between the NOE and NOE/T-ROE data sets are diminishing as illustrated in Figure 3.37. Similarly the difference is minute in population determined from six or seven structures.

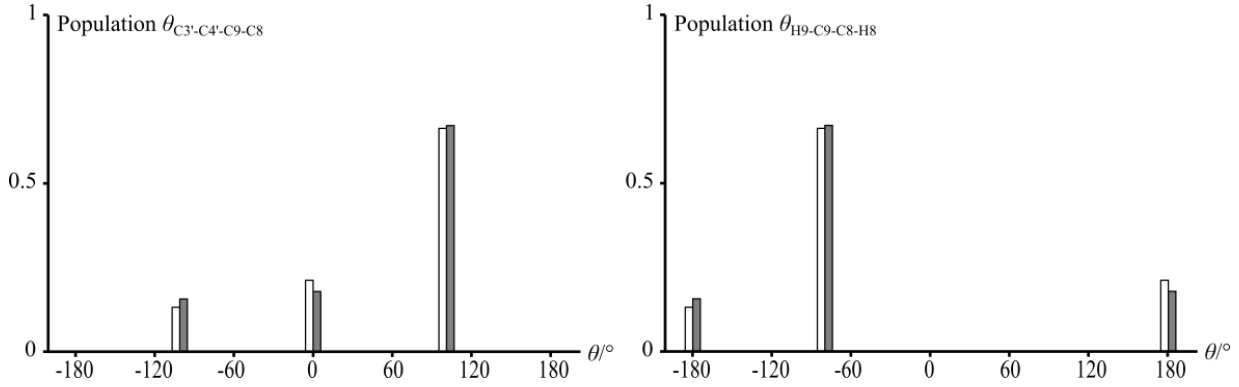



**Figure 3.37. The populations of quinine conformers from NOE (white) and NOE/T-ROE (grey) distances using the three essential conformers identified in Section 3.4.1.**

### 3.5.3 Conclusion and problems in including correlation times

While it would seem that including more data is always beneficial, the problem with the rotational correlation time is the erroneous methods that may be used to determine $\tau_c$. The robustness of the NOE methods, due to the $r^{-6}$ relationship, is thus reduced by introducing the correlation time, which does not benefit from a similar relationship. Due to generally low rotational correlation times in especially chloroform, any errors in NOESY and/or ROESY data will lead to large errors in the experimentally determined $\tau_c$, as seen Figure 3.38, due to only minor differences between the observables. As a consequence, an acceptable experimental error in the relative PANIC intensity of one to two will often be larger than the expected differences in PANIC intensities from NOESY/ROESY or two-field analysis experiments. For longer distances especially the introduction of the rotational correlation time into the equations is problematic, due to a larger influence of errors on the resulting distances.

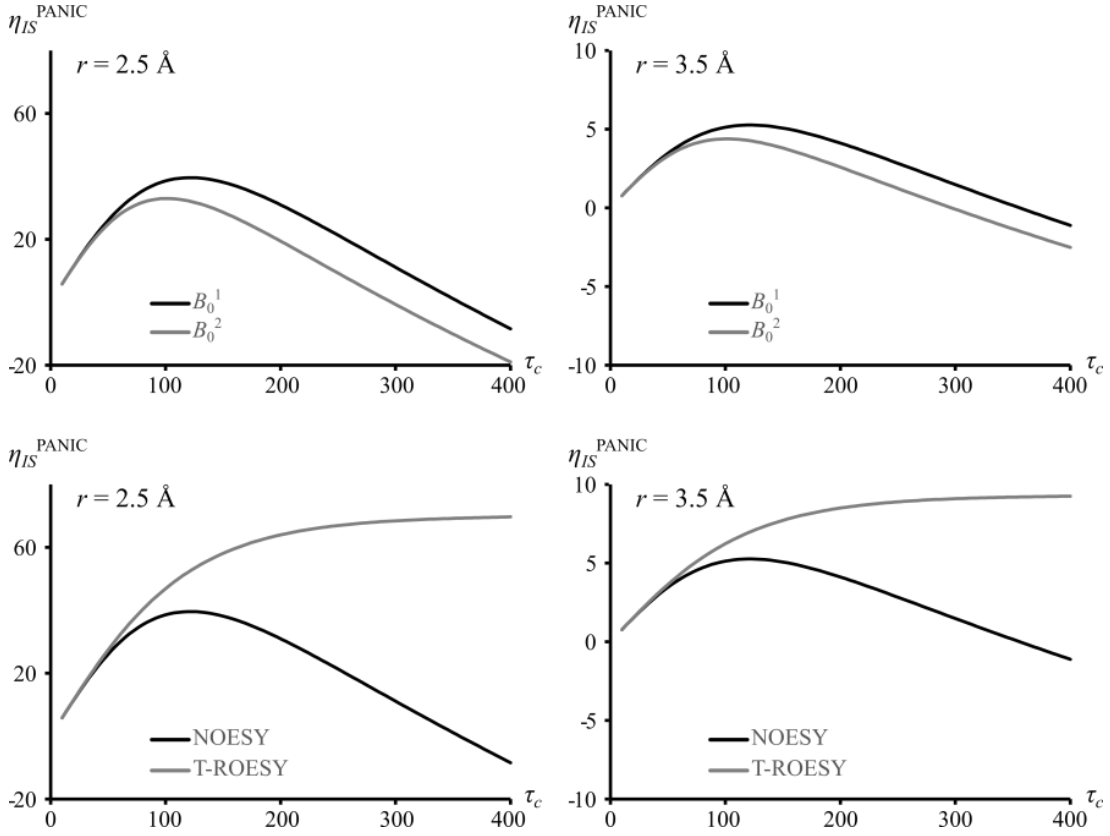



**Figure 3.38. Theoretical PANIC intensities from NOESY and T-ROESY or two-field NOESY analyses. Calculated from (3.8) and (3.11) using $\omega$ = 500 MHz (NOE/T-ROE), or $\omega_1$ = 500 MHz ($B_0^1$) and $\omega_2$ = 600 MHz ($B_0^2$). *r* given in the figures. A spinlock field of 5452 Hz with an offset at 5 ppm was assumed for the T-ROESY calculations.**

The promise of more exact distances is present from these types of analyses. As seen in Figure 3.39 the influence on the rotational correlation time on the distances obtained may be rather large, if the correlation time is sufficiently different. It is an enigma though, as the biggest difference from small rotational correlation time differences is in the regions where $\tau_c$ is hardest to measure correctly. In most cases it is thus not preferable to introduce the corrections into the equations for small molecules as this will only lead to more errors. It is also quickly evident that care should be taken in the NOESY crossover region, and that the ROESY experiment is not just superior in the relative

intensity, but also in a minimal dependence on the rotational correlation time in the crossover region (these two observations are coupled but maybe not always realized). Of course ROESY type experiments have their own challenges in this regard and may be in need of correcting before being used quantitatively.

In conclusion the attempt to obtain experimental distances with a higher correlation to those of 3D structures was not successful. While the methods may be used to estimate $\tau_c$ and check the assumption of roughly equal $\tau_c$ between proton pairs, direct usage of $\tau_c$ in distance calculations is not recommended.

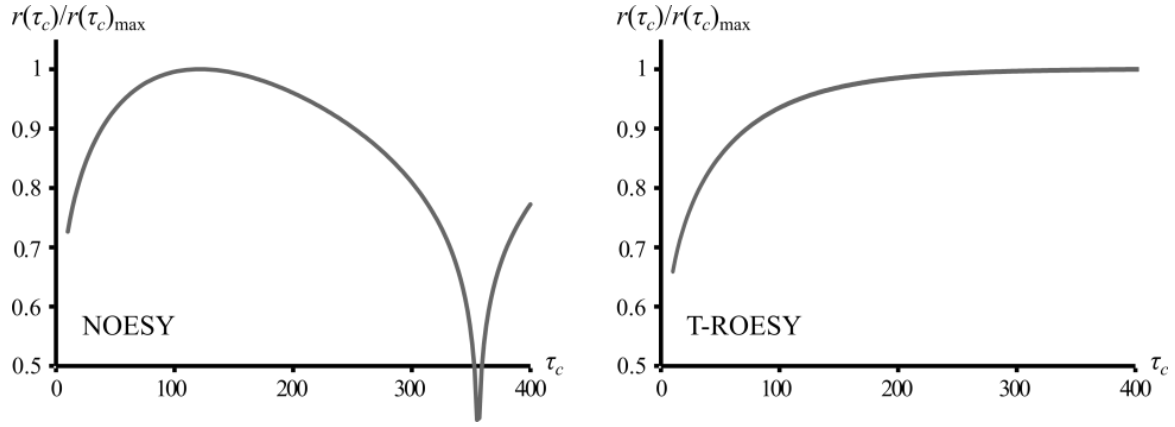



**Figure 3.39. Theoretical distance correction factors as a fucntion of the rotational correlation time (in ms) for NOESY and T-ROESY. Calculated from (3.8) and (3.11) using $\omega$ = 500 MHz, and a spinlock field of 5452 Hz with an offset at 5 ppm.**

## 3.6 Experimental

**General on structural calculations**
The theory used for obtaining distances from NOESY and ROESY experiments is outlined in the theory section (Section 2.1). A build up curve was always established using (25), 50, 100, 150 and 200 (and 300) ms mixing times for each cross-peak and each curve were evaluated. Standard setups for all NMR experiments and the spectrometers used are found in Appendix A1.

All *J*-coupling constants were extracted from 1D $^{1}H$ spectra, or DQF-COSY spectra if necessary due to overlapping resonances.

**Fungal extraction**

**Solvents used**
The solvents used were all HPLC grade. The water used was milliQ water from the milliQ Water Purification system by Merck Millipore.

**Media used**
The media used was yeast extract sucrose agar (YES) obtained from the stock at the Department of Systems Biology.

**Fungi**
*Aspergillus homomorphus,* IBT 21893 (or CBS 101889), originating from soil in the area of the Dead Sea in Israel. Inoculated by three point stabs. All plates were kept at 30 °C in darkness after inoculation for 7 days.

**Extraction**
Fungi were collected in double bagged stomacher bags; ten plates to a bag. 150 mL EtOAc (1 % FA) was added and the content processed in a Stomacher for 30 to 60 seconds. After one hour the extract was decanted into vials or flasks through a filter, the solvent was removed using $N_2$ and rotary evaporators. The bags were refilled with 150 mL EtOAc (1 % FA) and left over night before being re-collected as described.

The precipitate from the pooled extracts were dissolved in 200 mL 9:1 MeOH:$H_2O$ and extracted with 180 mL heptane. The MeOH/$H_2O$ phase was added 160 mL $H_2O$ and extracted 5 times with 100 mL DCM. All phases were dried using $N_2$ and rotary evaporators.

**Isolera One**
An Isolera One flash chromatography system by Biotage with auto-fractionation was used for fractioning. The column was a 120 g Biotage Snap KP-C18-HS with a column volume of 132 mL and fractions of 115 mL (max) were collected, with a flow speed was 40 mL/min.

**Waters**
A semi-preparative HPLC was used for further purification; the Waters 600 Controller coupled to a Waters 996 Photodiode Array Detector by Waters with a flow rate of 5 mL/min, using a Luna II

$C_{18}$, 5 μm, 250 × 10 mm column by Phenomenex. The runs are found Table 3.37. The samples were dissolved in MeOH before injection.

**Table 3.37. Waters runs. a: + 50 ppm TFA. b: isocratic run.**

| Run Fraction | Column | Gradient MeCN[a]/$H_2O$[a] | Run time [min] | Fraction name | Name | Amount isolated |
|---|---|---|---|---|---|---|
| CHO012206 | Luna II | 30-40 % MeCN | 17 | CHO013409 | CM-B | 1.1 mg |
| CHO013406 | Luna II | 20-60 % MeCN | 20 | CHO013801 | CM-A | 2.3 mg |

**NMR – natural compounds**

All compounds were dissolved in DMSO-$d_6$ in 3 mm (180 μL) or 5 mm (500 μL) NMR tubes. Standard pulse sequences used for structural elucidation, at either 500 or 800 MHz as described in Appendix A1. 2D ROESY spectra were used for distances calculations of homomorphosin A (23 mM) using $\tau_m$ = 200 ms and 2D NOESY spectra for cyclomorphosin A (15 mM), $\tau_m$ = 200 ms. For epi-10,23-dihydro-24,25-dehydroaflavine (43 mM) 2D NOESY spectra were used, $\tau_m$ = 150 ms. The remaining natural products, homomorphosterol (12 mM), aculene A (19 mM), aculene B (23 mM) and aculene C (18 mM) were solved from 1D $^1$H and DQF-COSY spectra.

**Simulations**

The modelling suite Maestro version 10.2.010 (2015) by Schrödinger was used for force field calculations,[126] using the program MacroModel version 10.8.[127,128] The MMFFs force field was used. To generate structures which should cover the conformational space of compounds, a conformational search was performed for each structure by the program MacroModel using torsional sampling with an energy cutoff of 50 kJ/mol, 100,000 steps and CPRG minimization.[127,129] The method will alter the dihedral angle of bonded nuclei by random Monte Carlo simulations, before it minimizes the new structure and, if it is indeed new as defined by relative atom position and within a given energy of the minimum, saves the structure. Also low-mode was used in certain simulations when stated in the text. Low mode sampling is based on an assumption that low-frequency vibrational mode eigenvectors connect energy minima by low-energy paths on the conformational energy surface. Thus by moving coordinates along a path given by the low-frequency modes, large energy barriers between energy minima may be traversed. These minima might not be identified by other methods due the large energy barriers connecting the minima. If moving along the low-mode eigenvector result in a new energy well, energy minimization is performed as for any other conformational sampling.[130]

Gaussian version 09 revision B.01 was used for DFT calculations including optimizations and NMR calculations.[131] Structures were optimized to a B3LYP/6-31(d) level of theory unless otherwise stated. NMR shielding tensors were calculated by MPW1PW91/6-311+G(2d,p) using GIAO and the PCM-SCRF model for implicit solvent.[47]

$J_{HH}$ were calculated by B3LYP/6-31g(d,p) u+1s considering only the Fermi Contact term. This is practically done by inserting the following in the Gaussian input file: nmr=fconly iop(3/10=1100000)

**Rigid to semi-rigid**

For rigid compounds which may be described by one to ten distinct conformers, e.g. quinine, DFT optimized structures are used as the basis of distance and *J*-coupling constant calculations. This is feasible due to the small amount of conformers, and the higher level of theory used should lead to more appropriate bond lengths, angles etc. This is equal to the approach utilized with much success in the group of Craig Butts, whom I visited on my external stay.[21,28–30] In general the populations of the conformers may here be described in detail from an iterative fitting of data from conformer populations to the experimental data. If multiple structures were used (e.g. quinine) the Excel Solver was used for conformer population analysis, using the evolutionary solving method. The reference distance was here allowed to differ by up to 5 % to fit the data the best, except for the one structure data of homomorphosin A and Epi-10,23-dihydro-24,25-dehydroaflavine.

**Flexible**

For more flexible compounds, where the amount of conformers is large and/or ill described, an approach based in DFT calculations is not feasible. One could use centroids and optimize these, but a lot of subtleties are lost, and the choice of centroids will introduce a bias into the data. The approach used was taken out of necessity, but the robustness of the approach makes it highly favorable. It was realized that the systems could not be described from a few representative conformers, and the approach instead compares the NOE/*J* data to a large amount of conformers from MD or MM simulations. The conformers are fitted iteratively to give the best correlation between data from the averaged conformers to the experimental data with no regards to the relative energy or other properties of the conformers. This means that the output conformations should be seen as representative of the conformational space inhabited by the compound. While the approach will by design be less discriminative in e.g. determining relative stereochemistry, due to the disregard of relative energy, it is less prone to misinterpretation due to faulty energy determination. Like all approaches it does rely on the different force field methods to output meaningful conformers, but by disregarding the energy of these, it is more likely that conformational space is covered. The cutoff of 50 kJ/mol was chosen for standard MM simulations to hopefully cover all of conformational space, and the NMR data was then used gain information of the actual conformational space. The high cutoff is used to diminish the influence of errors in the calculations. Since the NOE data is dominated by short distances, the average distances observed may be changed dramatically by short distances in lowly populated conformers, and thus excluding these may severely influence the conclusions.

The comparison of experimental data to that of 3D structures was performed by an in house written Matlab® script which iteratively minimizes the MAE using 10,000 steps. In the structural evaluation and iterative identification of the structures which gave the highest correlation to the NOE and *J*-coupling constants, both datasets were fitted independently and mixed, by weighing the differences as $\Delta=\Delta(\mathrm{NOE})+1/10\Delta(J)$. The latter data is shown throughout. Thus the unconstrained NMR structures were determined by averaging only the conformations that contributed to a best overall correlation to the data, by the lowest RMSD in both distances and coupling constants. The reference distance was here always equal to that of the averaged structures. The averaging was performed as described in Section 2.1.9 and 2.2.1.

**Marfey's reagent**

100 µg of homomorphosin A, cyclomorphosin A or aculene A and B were hydrolyzed with 200 µL 6 M HCl in a 2 mL vial at 110 °C for approximately 24 h. The solvent was removed at $N_2$ evaporator. To the vials, and vials with 50 µL (2.5 µmol) L- or D-amino acids (valine or proline), were added 50 µL $H_2O$, 20 µL 1 M $NaHCO_3$ and 100 µL 1 % 1-fluoro-2,4-dinitrophenyl-5-L-alanineamide (FDAA) in acetone. The vials were left for 40 °C for 1 h. The solvent was removed at $N_2$ evaporator, MeOH added and HPLC-DAD-MS data acquired.

**Synthetic compounds**

The synthetic compounds were supplied by the group of Prof. Christian A. Olsen (azumamides) or Prof. Morten Meldal (molecular recognition)

**Azumamides**

The azumamide analogues were dissolved in DMSO-$d_6$ in 3 mm tubes (180 µL). Spectra were acquired at 500 MHz and 800 MHz using 2D NOESY (500 and 800 MHz) and 2D ROESY (500 MHz) sequences, $\tau_m$ = 150 ms. 2D ROESY spectra at 500 MHz were used due to a superior *S*/*N*. 4096 data points were recorded in the direct and 512 in the indirect dimension. Relaxation delays were from 2 to 4 s, tested up to 7 s (>5 × $T_1$). Most nuclei were determined to have $T_1$ values up to approximately 1 s to 1.3 s relaxation times. All *J*-couplings were extracted from 1D $^1$H or DQF-COSY spectra, and *J*-couplings from dihedral angles were calculated by the Karplus equation or the HLA equation for the *β*-peptide protons.

Constrained simulations in implicit DMSO were conducted in MacroModel using the force field MMFFs or in implicit water using OPLS2005. DMSO was treated as a dielectric constant of 47.0. A conformational search was performed by the program MacroModel using torsional/low-mode sampling with an energy cutoff of 30 kJ/mol, 30,000 steps and CPRG minimization.[127,129] The distances from NOE intensities were applied to the structure and allowed to differ 20% governed by a force constant of 100 kJ/Å2.

Molecular dynamics simulations of azumamide A, desmethyl-azumamide C and epi-azumamide E were performed by the program Desmond, carried out by Post. Doc. Niels Christensen.[128] Individual structures were solvated in a TIP3P or DMSO box using a 10Å buffer, and neutralized by a single $Na^+$ counter ion. The default minimization protocol in Desmond was used, consisting of minimization with solute constraints, minimization without constraints, Berendsen NVT simulation at T = 10 K with small time steps and constraints on heavy solute atoms, Berendsen NPT simulation at T = 10 K with constraints on heavy solute atoms, Berendsen NPT simulation with constraints on heavy solute atoms, and unrestrained Berendsen NPT simulation. In the MD production runs, 250 ns NPT simulations with periodic boundary conditions were run for each system. The Nose-Hoover, 8 chain thermostat was used to regulate temperature to 300 K with a relaxation time of 1.0 ps. The Martyna-Tobias-Klein barostat was used to regulate pressure to 1 bar with isotropic coupling and a relaxation time of 2.0 ps. Equations of motion were integrated using the RESPA10 integrator with bonded (2.0 fs), near (2.0 fs), and far (6.0 fs) time steps. Non-bonded interactions were subjected to a 9 Å cut-off. Long-range electrostatics was treated with the smooth-particle mesh Ewald method . All atomic coordinates for the simulations were saved at 10 ps intervals.

**Molecular recognition**

The Target 20 (33 mM) and Ligand 7 (34 mM) were dissolved in $H_2O$/10% $D_2O$ in 5 mm tubes (0.5 mL). The pH was adjusted by NaOH (0.1 M). Spectra were acquired at 800 MHz using 2D NOESY and 2D ROESY sequences using water suppression using either presaturation or watergate. Distances obtained from both types of water suppressions were compared and the differences were negligible. 2D NOESY spectra using presaturation were used, $\tau_m$ = 200 ms. Prior to acquisition $N_2$ was bubbled through the sample to remove $O_2$. The solutions were mixed, 0.5 mL transferred to a NMR tube and the acquisitions repeated as for the isolated peptides.

Titrations of ligand 7 with target 20 were performed at pH 5.5 and 6.5. Both solutions of the peptides were adjusted to the target pH prior to titrations, which were performed in 3 mm NMR tubes. 0.9 (pH 5.5) or 0.7 (pH 6.5) mg of Ligand 7 was dissolved in 180 μL 10% $D_2O/H_2O$ and adjusted to the target pH with 0.1 M NaOH and double the molar amount of Target 20 was dissolved in 100 μL 10% $D_2O/H_2O$ and pH adjusted with 0.1 M NaOH. Titrations were then performed stepwise by adding the amount stated (e.g. 10 % equals to 4.8 μL Target solution added) in the figures, and 1D $^1H$ spectra were acquired with 128 scans.

The structures, Target 20 and Ligand 7, were built in MOE and both subjected to constrained dynamics in water from NOE derived distances by Prof. Morten Meldal. Starting from several rounds of 3 ns annealing from 700 to 300 K, using AMBER 12 force field, the resulting structures were simulated at 300 K for 5 ns and energy minimized. The NOE-constraints from the complex were used for the initially as the set of distances were larger and presented better *S*/*N* than those of the individual compounds. The individual structures were manually interacted assuming interactions of arginines and carboxylates, and the resulting complexes were subjected to molecular dynamics. Two orientations, out of six attempted, gave no significant violation of NOE's or change of structure. These two orientations were subjected to extended molecular dynamics for 5 ns with the constraints maintained. One orientations lead to productive binding and surface complementarity and gave a significantly lower calculated energy compared to the other. In this structure all backbone distances constraints determined from relative NOE intensities were accommodated. The complex was optimized again without NOE constraints, and only minor adjustments of the two structures were observed.

NOE constraints from for individual Target 20 and Ligand 7 experiments, was employed to structures from the optimized complex, which were then subjected to annealing from 700 to 300 K. The Target 20 was essentially unchanged the Ligand 7 changed to a partially opened.

**Other compounds**

Spectra of strychnine (124 mM), progesterone (127 mM) and quinine (123 mM) were all acquired in 5 mm tubes (0.5 mL), dissolved in the solvents given in the text (concentration in $CDCl_3$, similar in DMSO-$d_6$). 1D NOESY and 1D ROESY experiments were used and only resonances with good baseline separation were used as irradiated peaks, $\tau_m$ = 300 ms. For the ROESY experiments a spinlock field of 5452 Hz with an offset at 5 ppm was used at 500 MHz. For two-field analyses the other field was 600 MHz. All structures were optimized to a MPW1PW91/6-311+g(d,p) level of theory.

# 4 Development of NMR experiments for determination of long-range *J*-coupling constants

*The experimental data presented in this chapter concerns the work on novel NMR experiments to extract coupling constants from small organic molecules, and is based on the following two articles. Work on the first article was performed in collaboration with Ole W. Sørensen, Louise Kjaerulff and Andrew J. Benie, and work for the second article was performed in collaboration with Ole W. Sørensen. Louise Kjærulff presented the work of combining the five pulse sequences in the original $S^3$ HMBC, vide infra, in details in her PhD thesis "NMR structural studies of oligosaccharides and other natural products",[132] and the topics will be mostly outlined in this text.*

Appendix A11: $S^3$ HMBC: Spin-State-Selective HMBC for accurate measurement of homonuclear coupling constants. Application to strychnine yielding thirteen hitherto unreported $J_{HH}$

Appendix A11: $S^3$ HMBC *hetero*: Spin-State-Selective HMBC for accurate measurement of long-range *heteronuclear* coupling constants.

## 4.1 3D structural information from long-range *J*-coupling constants

The most commonly used 3D structural information obtained from *J*-coupling constants is based on data from $^3J_{HH}$, which was discussed in Section 2.2.1 where general scalar coupling theory is also found. The focus on $^3J_{HH}$-coupling constants is mainly due to in the ease of extraction of the coupling constants and the availability of a plethora of Karplus- or Karplus-like equations, which makes dihedral information readily obtainable.[40,42,133] While the more generalized Karplus-like equations may be less accurate than e.g. DFT, the ease of calculation is a major benefit, and the scope of use is continually extended by specific equations for specific tasks.[134]

For long-range *J*-coupling constants, defined as $^nJ_{XH}$, with n > 3 for $^1$H and n > 1 for $^{13}$C, the coupling constants are not generally easily obtained nor easily correlated to a 3D structural parameter. While some Karplus equations have been developed for $^2J_{CH}$ and $^3J_{CH}$ systems, they are generally very specific and used in only a few fields, perhaps due to difficulties in accessing the coupling constants or lack of general applicability of the equations.[46,135–137] It has been suggested to utilize a strong/medium/weak coupling constant approach, in spiritual resemblance to NOE work for macromolecules.[46] This approach is a viable option for rigid structures, but difficulties may be envisioned from an increase in flexibility, even though some examples of flexible structures were presented.[46] For other long-range coupling constants ($^nJ_{XH}$, n > 3, X=H,C) Karplus equations are complicated by addition of extra bond and thus extra dihedral angles to the equations, though very specific equations were empirically derived for specific cases of $^4J_{HH}$-coupling constants.[138,139]

The data from the long-range coupling constants is thus usually compared to more time-consuming theoretical DFT calculations to gain structural information.[53,140,141] While this may limit the amount of conformers included in theoretical data calculations, access to increasingly larger computers decreases the calculation time and has made the information increasingly accessible. Since any increase in information of structural parameters will inevitably lead to more structural knowledge, the information from long-range coupling constants holds considerable value and should not be discarded.[6]

### 4.1.1 Published methods to extract long-range coupling constants

This section will serve as a brief introduction to the currently available NMR experiments which may be used the extraction of long-range coupling constants. Regarding the extraction of long-range homonuclear coupling constants, E.COSY type experiments or standard 1D $^{1}$H or DQF-COSY sequences may be used by simulation of coupling patterns or displacement of resonances.[142–144] In the DQF-COSY case the measurements described were performed on rigid structures with larger long-range coupling constants.[144]

In comparison, a plethora of experiments has been developed for the extraction of heteronuclear coupling constants, but the NMR community has still to decide on a golden standard, akin to the 1D $^{1}$H spectra for homonuclear coupling constants.[145,146] Characteristics for a golden standard is, per Parella and Espinosa,[145] an experiment which displays good sensitivity and high resolution, is generally applicable, accurate, sign-selective and simple to run and analyze. This experiment could of course be the coupled 1D $^{13}$C spectrum, but here the analysis is cumbersome and NMR simulations are usually necessary due to the complexity of the multiplets.[147–149] The lack of a reference experiment, in addition to the differences between heteronuclear coupling constants determined from the available experiments, may in principle lead to overfitting of data. This will be discussed *vide infra*.

The experiments for determining long-range heteronuclear coupling constants may be, coarsely, divided into two categories, as being either COSY/TOCSY or purely HMBC/HSQMBC based.[145] Note that the experiments in TOCSY category of course also utilize long-range heteronuclear coupling constants and may thus also be perceived as e.g. HSQMBC based. A schematic representation between the TOCSY and HMBC/HSQMBC based coupling topologies is found in Figure 4.1.

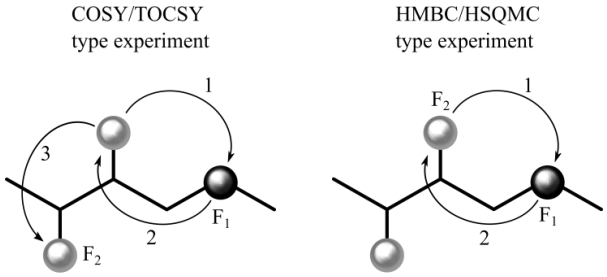



**Figure 4.1. The *J*-coupling transfer mechanism leading to spin correlations of COSY/TOCSY and HMBC/HSQMBC type experiments. The observed coupling constants are indicated by the observed nuclei in the F1 and F2 dimension. The COSY/TOCSY type utilizes an $^{n}J_{CH}$-coupling to build a correlation which is propagated to neighboring protons from $^{n}J_{HH}$-coupling. The HMBC/HSQMBC type utilizes only the $^{n}J_{CH}$-coupling. Grey: $^{1}$H, black $^{13}$C. Inspired from literature[145]**

The NMR experiments currently present vary with respect to how the *J*-coupling constants are extracted, from *J*-quantitative (comparison of cross-peak volume) to peak-splittings measured as in- or antiphase (IP or AP), IPAP, *J*-resolved, E.COSY type and TROSY patterns, exemplified in Figure 4.2.[145]

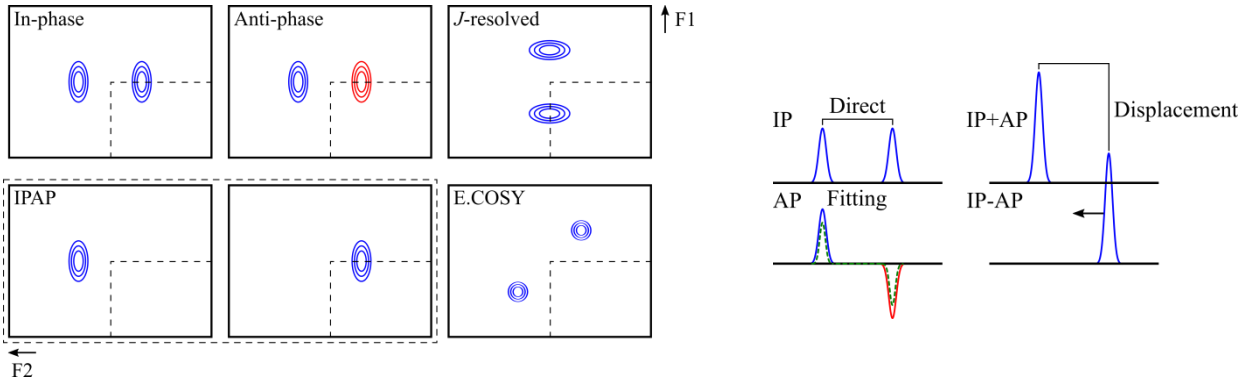



**Figure 4.2. Illustration of cross-peaks and cross sections from the different methods for the extraction of coupling constants. Inspired from literature.[145]**

The determination of the coupling constants from the coupling patterns, range from integration, direct measurements of peak-splitting and displacement of subspectra to more or less complicated theoretical simulations of multiplets.[145] For experiments where the *J*-coupling constants are measured in the F1 (indirect) dimension, the resolution is determined by number of increments and spectral width, which will often lead to excessively long experiment times. To alleviate this problem, the *J*-coupling constants are usually scaled by a factor *k*, which may be chosen according to the minimum required discrimination of observable *J*-coupling constants. For example; if coupling constants of approximately 1 Hz need to be determined and $k = 15$, the FID resolution in F1 needs to be 15 Hz per point. The size of *k* is restricted by the fact that the coupling evolution time increases with *k*, and thus large values may be problematic due to $T_2$ relaxation of the nuclei and concurrent losses in sensitivity.

Some of the published experiments are given in Table 4.1 to exemplify the peak patterns, measurement method, and general considerations. Advantages and disadvantages of the presented experiments will differ depending of the information needed, and are to some extent subjective.

**Table 4.1. Examples of experiments for the determination of long-range coupling constants. Peak pattern (see Figure 4.2 for examples), type and references are given. Note that several experiments may also be used for determination of $^{n}J_{NH}$ as well as $^{n}J_{CH}$-coupling constants. The table is inspired by Parella *et al.*[145]**

| Name | Peak pattern | Type | Comments | Ref. |
|---|---|---|---|---|
| Quantitative HMBC | *J*-quantitative | $^{n}J_{CH}$ | Compares the HMBC peaks by volume and calculation is needed.<br>Not sign-selective. | 150,151 |
| EXSIDE | *J*-resolved | $^{n}J_{CH}$ | Not sign-selective.<br>F1 dimension limit resolution.<br>Selective version exists to e.g. increase resolution. | 152,153 |
| XLOC | E.COSY | $^{n}J_{HH}$ | HH in F2 dimension.<br>Not sign-selective. | 154 |
| *J*-HMBC | *J*-resolved | $^{n}J_{CH}$ | Not sign-selective.<br>F1 dimension limit resolution. *J*-scaling is used. | 154 |
| HECADE | E.COSY | $^{n}J_{CH}$ | CH in F2 dimension.<br>Sign-selective | 155 |
| P.E.HSQMBC | E.COSY | $^{n}J_{CH}$<br>$^{n}J_{HH}$ | CH and HH coupling constants.<br>Sign-selective for HH. Not sign-selective for CH.<br>Complicated patterns.<br>CH in F1 dimension. *J*-scaling is used. | 156 |
| HSQC-TOCSY | In-phase/IPAP | $^{n}J_{CH}$ | CH in F2 dimension.<br>Sign-selective version exists. | 157–159 |
| HSQMBC | In-phase/ anti-phase /IPAP | $^{n}J_{CH}$ | CH in F2 dimension.<br>Not sign-selective. | 160,161 |
| Phase sensitive HMBC | Anti-phase | $^{n}J_{CH}$ | CH in F2 dimension.<br>Not sign-selective.<br>Simulations of cross-peaks needed. | 162 |
| HSQMBC-COSY/TOCSY | IPAP | $^{n}J_{CH}$ | Sign-selective.<br>CH in F2 dimension.<br>Selective versions are available. | 163 |

The presented list is not complete, nor are the given comments, and multiple experiments not mentioned here are available in the literature and e.g. various reviews include additional examples.[145,164] In the following sections, NMR experiments that provide easy access to $^{n}J_{HH}$ as well as $^{n}J_{CH}$ including the relative sign of the coupling constants will be described.

## 4.2 $S^3$ HMBC *homo*

The spin-state selective ($S^3$) HMBC *homo* (homonuclear in regards to the determined *J*-coupling constants) pulse sequence was developed in order to measure long-range homonuclear coupling constants utilizing the $S^3$ or E.COSY methodology of subspectra displacement.[165,166] The practical determination of coupling constants is thus in line with the IPAP method described in the previous section.[145] The subspectra are linear combinations of spectra of a standard HMBC pulse sequence and a second pulse sequence with an additional $\pi/2$ $S_z$ rotation under the one bond coupling evolution. The selected multiplicity edited pulse sequences are the improved multiplicity edited HMBC as the standard (a-c) and the HAT HMBC as the second (d,e) pulse sequence, see Figure 4.5 and equation (4.1).[167–170] The experiment was thus written as a pseudo 3D experiment, running five 2D experiments from the five pulse sequences found in Figure 4.5. The linear combinations of the five resulting spectra that give rise to the $S^3$ HMBC subspectra are given by equation (4.1).

$$\underbrace{\underbrace{\left\{a \pm \tfrac{1}{2}[b+c]\right\}}_{\text{odd/even}} \pm \underbrace{\{d \pm e\}}_{\text{odd/even}}}_{S^3} \qquad \textbf{(4.1)}$$

As indicated in equation (4.1) the terms $a \pm \frac{1}{2}[b+c]$ and $d \pm e$ are used to achieve multiplicity editing, while inter-brace addition/subtraction leads to $S^3$ editing of the active coupling constants. The active coupling is here defined comparable to DQF-COSY, where active couplings are anti-phase and passive in-phase, as the active coupling is observed between subspectra and the passive couplings are observed in a sub-spectrum, sometimes in the form of line-broadening. In the $S^3$ HMBC *homo* the active coupling is thus the long-range proton-proton coupling constants and the passive couplings are the long-range heteronuclear coupling constant and possibly coupling constants to other protons.

The spectra could in principle be combined in eight different manners but odd and even data are not mixed in practice, and the four are used where the sign in odd/even braces matches in equation (4.1). For clarification the spectrum with e.g. methine/methyl carbons from (a-½[b+c]+[d-e]) will be referred to as “add” or “+” and (a-½[b+c]-[d-e]) as “subtract” or “-“ in this text.

The result is eight distinct HMBC spectra as given in Figure 4.3, where four (5-8) are used for coupling constant extraction, and the other four may be used as standard multiplicity edited improved or HAT HMBC spectra. The sorting of the spectra is performed automatically by an in house developed AU-program which works seamlessly with Topspin, and the numbers in Figure 4.3 correlate to spectrum number of from the script.

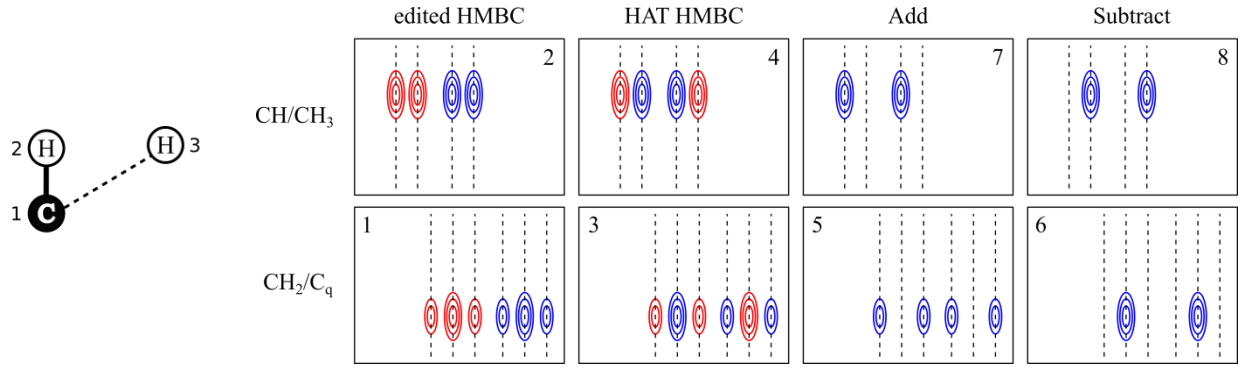



**Figure 4.3. The eight (numbered) output spectra from the $S^3$ HMBC *homo* experiment as given by the AU program. The chemical shift of carbon 1 is along F1, the chemical shift of proton 3 along F2 and the $S^3$ edited coupling constant of protons 2 and 3 is the displacement between the resonances in the add/subtract spectra. It is assumed that $J_{CH} > J_{HH}$ and that the multiplets shown are a CH···H(singlet) and a $CH_2$···H(singlet).**

The resulting subspectra are used to extract the long-range coupling constants including the value and relative sign. The coupling constants are determined from displacement (or direct measurements) of the peaks in matching +/- subspectra. The extracted coupling constant is between proton 2 in Figure 4.3, attached to the carbon with chemical shift in F1, and the observed proton 3 with chemical shift in F2. The proton 3 and carbon 1 are coupled from the long-range heteronuclear coupling constant between the two, which is observed as a passive splitting in the spectra. A schematic representation of the $S^3$ editing is indicated in Figure 4.3 and further explained in Figure 4.4.

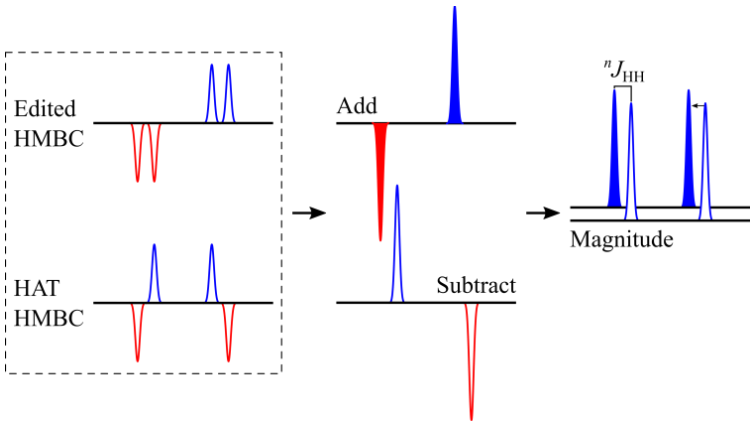



**Figure 4.4. Principle of the $S^3$ editing of CH--H in the $S^3$ HMBC experiment. Addition and subtraction of the edited- and HAT HMBC spectra leads to subspectra which, in magnitude mode, may be used to extract the $^nJ_{HH}$-coupling constant. The second coupling constant present in the spectra is the $^{n-1}J_{CH}$-coupling constant. It is assumed that $J_{CH} > J_{HH}$ and that the multiplet is a CH···H(singlet).**

The absolute sign may be determined by keeping track of the $S^3$ add- and subtract spectra. The coupling is positive when the obtained subtract spectrum is positioned downfield relative to the add spectrum, if it is upfield the coupling constant is negative. This is governed by the sign of the $^1J_{CH}$-coupling constant which is always positive. For practicalities; if the subtract spectrum is used for displacement in Topspin, the size and sign of the coupling constants are extracted directly as presented in the program.

The five pulse sequences were used as previously published, with minor modifications to the HAT sequences.[167–170] In order to ensure equal length between the five sequences, delays were added. An initial delay of $\tau + \varepsilon'/2$ was added before, and $\tau/2$ just after, the first $^{1}$H $\pi/2$ pulse of the HAT experiments, see the Figure 4.5 caption for a definition of the delays. A decoupling period of $\tau + \varepsilon'/2 +$ $^{1}$H $\pi/2$ was also implemented in accordance with the three standard HMBC experiments. Lastly, a delay of $\tau/2$ was added after the last $^{13}$C $\pi/2$ pulse. This led to pulse sequences which uphold basic conditions for combinatory pulse sequences, as the time from first $^{1}$H $\pi/2$ pulse to FID and the chemical shift and coupling constant evolution match between all five sequences. The magnetization thus experiences identical times of relaxation in all five sequences. The pulse sequences were programmed to run interleaved, with [b] and [c] spectra alternating to satisfy equation (4.1) without an initial need for scaling and thus an optimized run time versus *S*/*N*.

In practice the spectra are usually scaled though, due to high editing accuracy of the edited HMBC and HAT HMBC experiments. By employing the linear combinations [b + c] and [d ± e] the magnetization components in the sequences that could give rise to out-of-phase error terms are diminished. This results in slightly lower intensities for these spectra compared to (a), except for cases of $CH_n$ pairs where $\tau$ in the sequences is exactly matched to $(2\ ^{1}J_{CH})^{-1}$. By scaling [b + c] by a constant $k$ and [d ± e] by $(2k/[k+1])^{½}$, the level of $J$ cross talk is reduced. This is hardly needed due to an already low level of cross talk, but slight improvements were obtained when using a small scaling constant (usually in the order of $k$ = 1.05). While it is possible to scale the cross-peaks individually to get optimal line shapes, this was not identified to be experimentally needed.

The add/subtract nature of the spectra is beneficial when considering the acquisition time, where four interleaved HMBC spectra may sound like a very long experiment to actually implement in normal acquisition setups. While a minimum of four scans are needed for each sequence, essentially setting a minimum time, the fact that the resonances are added to and subtracted from each other result in a gain of *S*/*N* in the final spectrum from every single scan. The broadband nature of the experiment and the fact that the coupling constants are measured in the F2 dimension also helps to keep the acquisition time short, due to only having to run one experiment per sample and the inherent high resolution.

For the comparison of long-range coupling constants, especially those of very low absolute values, to theoretical values from 3D structures, a problem has often been obtaining a high degree of accuracy. Thus dividing the coupling constants into categories such as low, medium, or high has been practiced.[46] This is in general a viable route for rigid structures, while it is problematic in e.g. conformational population determination. The accuracy of the extracted coupling constants is thus of utmost importance. In the following, extracted coupling constants will be compared to theoretical coupling constants with a high degree of correlation, even for very small coupling constants. To ensure that the displacements were not reliant on the human eye of the spectroscopist, in this case the author, a Matlab script was developed which automatically identifies and extracts coupling constants from the spectra, see Appendix A6.

The pulse sequence was tested on multiple compounds, including strychnine and isopinocampheol (IPC), and selected results and points are included *vide infra*.

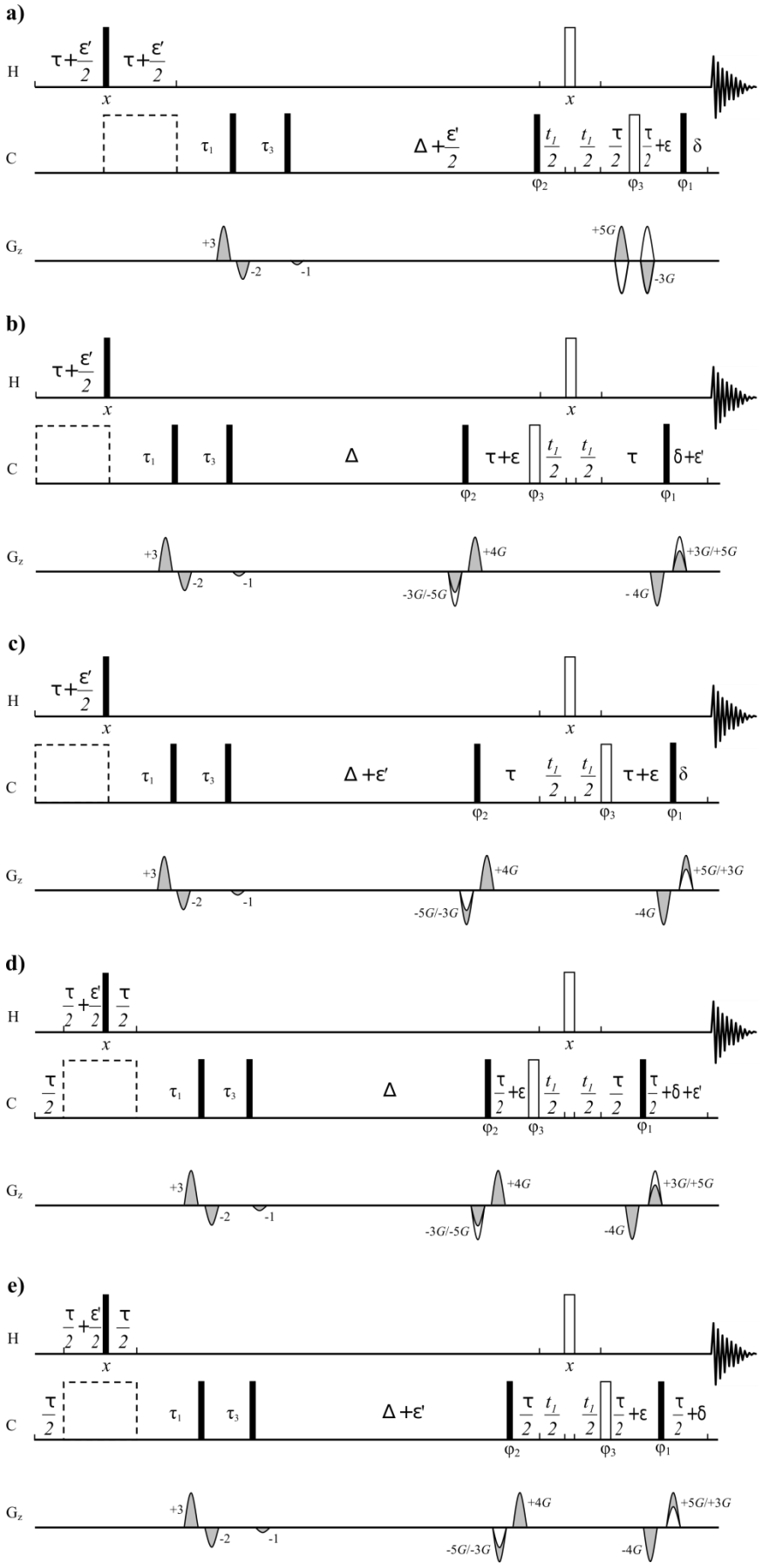



**Figure 4.5. S$^3$ HMBC *homo* pulse sequence comprised by a-c; the edited HMBC sequences, and d,e; the HAT HMBC sequences, all shown with a 2nd-order low-pass *J* filter (LPJF), which may be exchanged to 3rd-order LPJF or removed.[166]**

**Filled and open bars refer to $\pi/2$ and $\pi$ pulses, respectively, and the dashed open boxes represent $^{13}$C decoupling.**

**$\tau = (2\ ^1J_{CH})^{-1}$ or $(^1J_{max} + {}^1J_{min})^{-1}$**

**$\delta$ is a gradient delay**

**$\varepsilon = 2\ t_{1/2,min} + t(\pi_H)$**

**$\varepsilon' = \varepsilon + t(\pi_C)$**

**$\tau_1 = \frac{1}{2}[^1J_{min} + 0.146\ (^1J_{max} - {}^1J_{min})]^{-1}$**

**$\tau_3 = \frac{1}{2}[^1J_{max} - 0.146\ (^1J_{max} - {}^1J_{min})]^{-1}$**

**$\Delta$ is the delay for evolution under heteronuclear long-range couplings and is set to $(2{\cdot}8\ Hz)^{-1}$ as standard.**

**Phase cycling is performed as**

**$\varphi_1 = \{x, -x, -x, x\}$**

**$\varphi_2 = \{x, x, 4(-x), x, x\}$**

**$\varphi_3 = \{4(x), 4(y), 4(-x), 4(-y)\}$.**

### 4.2.1 Computation of long-range homonuclear coupling constants

The computation of homonuclear coupling constants may be achieved by two methods as previously discussed; Karplus type equations or density functional theory (DFT) methods.[40,42,53] No appropriate Karplus equations were identified for the long-range coupling constants and DFT methods were used. The taken approach is largely based on an article of Bally and Rablen, which establishes a good practice in calculating standard homonuclear coupling constants.[53] Since most studies of homonuclear coupling constants have focused on the easily accessible dihedral and germinal coupling constants, the applicability of methods and functionals to the long-range coupling constants needed to be evaluated, and $S^3$ HMBC *homo* data compared to multiple different functionals is found in Appendix A6.

In line with the results from Bally and Rablen, the optimal low-cost method of calculating homonuclear coupling constants was identified as optimization to B3LYP/6-31G(d), and calculation of the FC term of the coupling constants at a B3LYP/6-31G(d,p) level of theory. The function was augmented with four compact 1s functions to give more accurately calculate the FC term by using the input iop(3/10=1100000) in Gaussian.[53] The resulting *J*-coupling constants were scaled by 0.9117 in accordance with the literature.[53] The correlation between experimental and theoretical coupling constants is surprising in that the literature study included only few long-range coupling constants, and none as small as observed from $S^3$ HMBC *homo* – and still the fit is excellent.

### 4.2.2 Results

The $S^3$ HMBC *homo* experiment resulted in spectra from which the long-range coupling constants were easily extracted; see Figure 4.6 for examples and full spectrum in Appendix A6. Many $^3J_{HH}$-coupling constants were extracted, making the $S^3$ HMBC *homo* an excellent experiment for obtaining exact $^3J_{HH}$-coupling constants in spectra of compounds with a crowded proton dimension due to the extra resolution obtained by utilizing the carbon chemical shift range.

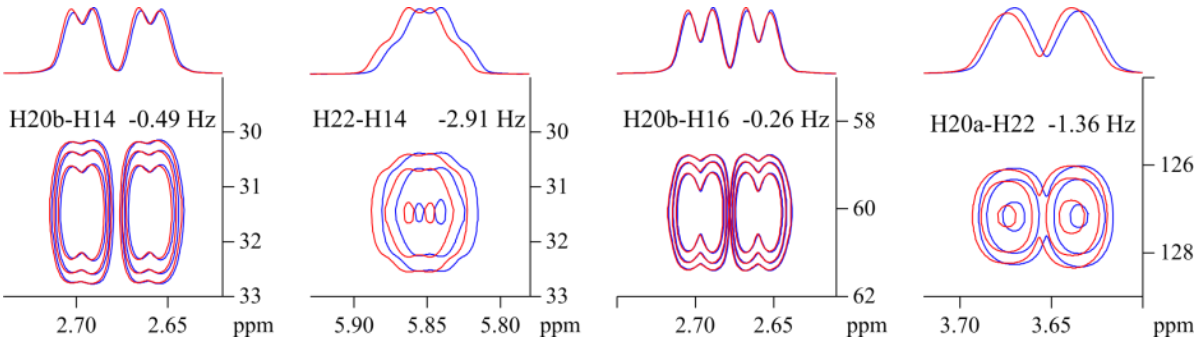
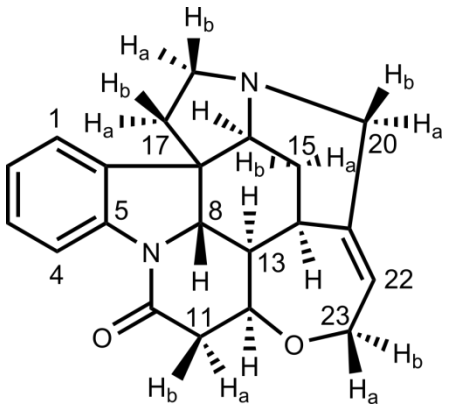


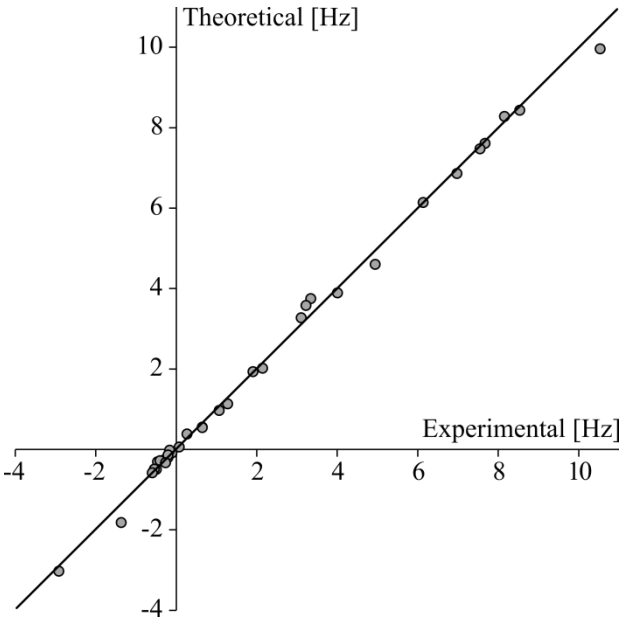

**Figure 4.6. Excerpts and cross-sections through F1 of cross-peaks of $S^3$ HMBC *homo* for strychnine (300 mM). Chosen as n > 3 for the associated $^nJ_{HH}$-coupling constants.**

A total of 34 coupling constants were determined for strychnine. As a testament to the novelty of extracting small long-range homonuclear coupling constants, 13 coupling constants were extracted from this much utilized model compound, which were not reported in the literature. The extracted long-range coupling constants of strychnine are compared to theoretical values in Table 4.2.

**Table 4.2. Left: Experimental and theoretical $^{n}J_{HH}$-coupling constants of strychnine measured at 400 MHz ([a]800 MHz). Bold indicates coupling constants only published from S$^3$ HMBC. Theoretical coupling constants calculated by B3LYP/6-31g(d,p) u+1s from B3LYP/6-31g(d) optimized structures. *The protons of 17ab overlap and a combined value was used. Right: Numbered structure of strychnine and experimental versus DFT coupling constants, where the line represents a perfect fit. Fit of data: a = 1.01, b = 0.00, $R^2$ = 0.997.**

| H1 | H2 | S$^3$ HMBC [Hz] | Theoretical [Hz] | n |
|---|---|---|---|---|
| 1 | 2 | 7.68 | 7.61 | 3 |
| 1 | 3 | 1.50, 1.28 | 1.13 | 4 |
| 1 | 4 | 0.65, 0.66 | 0.55 | 5 |
| 2 | 4 | 1.07, 1.24 | 0.97 | 4 |
| 2 | 3 | 7.55[a] | 7.47 | 3 |
| 3 | 4 | 8.16 | 8.28 | 3 |
| **8** | **12** | **-0.16, -0.01** | **-0.10** | **4** |
| 8 | 13 | 10.58, 10.54 | 9.96 | 3 |
| **8** | **16** | **-0.32, -0.20** | **-0.27** | **4** |
| 11a | 12 | 8.54 | 8.43 | 3 |
| **11a** | **13** | **-0.46** | **-0.31** | **4** |
| 11b | 12 | 3.35 | 3.75 | 3 |
| 12 | 13 | 3.23 | 3.58 | 3 |
| **12** | **14** | **-0.20** | **-0.20** | **4** |
| **12** | **15a** | **-0.13** | **-0.09** | **5** |
| **12** | **23a** | **-0.10** | **-0.09** | **4** |
| **12** | **23b** | **-0.16** | **-0.02** | **4** |
| 13 | 14 | 3.11 | 3.27 | 3 |
| 13 | 15a | 0.27 | 0.38 | 4 |
| **13** | **15b** | **-0.40** | **-0.28** | **4** |
| 14 | 15a | 4.95 | 4.60 | 3 |
| 14 | 15b | 1.91 | 1.93 | 3 |
| **14** | **16** | **0.65** | **0.54** | **4** |
| **14** | **20b** | **-0.49** | **-0.50** | **4** |
| 14 | 22 | -2.91 | -3.03 | 4 |
| 15a | 16 | 4.01 | 3.89 | 3 |
| 15b | 16 | 2.15 | 2.02 | 3 |
| 16 | 17a | - | 0.04 | - |
| 16 | 17b | - | -0.18 | - |
| **16** | **17ab*** | **-0.20** | **-0.14** | **4** |
| **16** | **18a** | **-0.54** | **-0.49** | **4** |
| **16** | **18b** | **0.08** | **0.06** | **4** |
| **16** | **20b** | **-0.26** | **-0.33** | **4** |
| 20a | 22 | -1.36 | -1.82 | 4 |
| 20b | 22 | -0.59 | -0.58 | 4 |
| 22 | 23a | 6.98 | 6.86 | 3 |
| 22 | 23b | 6.14 | 6.14 | 3 |




It was only possible to obtain one coupling constant for some methine to methine correlations due to spectral overlap. When two coupling constants were identified, they were generally in good correlation with a maximum deviation of 0.22 Hz. In principle it is possible to extract coupling constants to methylene protons from a multiplet containing both coupling constants. In practice, the coupling constants to the methylene protons need to be large and different in order to determine coupling constants with a good correlation to the methine data, shown in Section 4.4.4. The general correlation of the data was excellent with a RMSD of 0.21 Hz for the full data set and, rather surprisingly, 0.15 Hz if only long-range coupling constants were considered. This was in part due to the values being lower, and relative to the size of the extracted coupling constants the differences

were larger. It is also an indication of the high accuracy of the experiments though, and that the theoretical calculations that work well for $^{3}J_{HH}$ seemed to also correlate well to the long-range coupling constants. As indicated in Table 4.2, mostly 3- and 4-bond coupling constants were observed, and the novel coupling constants from the S$^3$ HMBC were all 4- or 5-bond and all below 1 Hz in magnitude.

## 4.3 S$^3$ HMBC *hetero*

An extended version of the homonuclear S$^3$ HMBC was developed to extract long-range heteronuclear coupling constants. As mentioned above, the heteronuclear coupling constants are generally less utilized than their homonuclear counterparts, in part due to difficulty of extraction. The potential for structural information it huge however, as the number of $^{2\text{-}4}J_{CH}$-coupling constants will be larger than the number of $^{2\text{-}5}J_{HH}$-coupling constants for the majority of organic compounds. An example could be strychnine, which is a proton rich compound with the molecular formula $C_{21}H_{22}N_2O_2$, which has 96 possible $^{2\text{-}5}J_{HH}$ and 198 possible $^{2\text{-}4}J_{CH}$-coupling constants. In practice only a fraction of these coupling constants is measureable due to e.g. low coupling constants leading to no signal or overlap of chemical shifts. For proton deficient compounds, which are usually hard to solve, the larger abundance of heteronuclear coupling constants is more pronounced. In determining heteronuclear coupling constants, the sign-selective nature of the S$^3$ HMBC is imperative as $^{2}J_{CH}$ ranges from approximately -6 to 8 Hz while $^{3}J_{CH}$ ranges from 0 to 9 Hz.[46]

### 4.3.1 Changes needed

In order to change the observed coupling constant from $^{n}J_{HH}$ to $^{n}J_{CH}$ a new pulse sequence element needed to be implemented. This element needed to be able to exchange the polarization from proton to carbon (or from βα to αβ) in e.g. $^{13}$C-$^{1}$H pairs, as illustrated in Figure 4.7, where a zero-quantum pulse transfers the polarization from spin 2 (C-H) to spin 1 (C-H) to access the long-range heteronuclear coupling constant. A similar approach was used in the original S$^3$ work.[171–174]

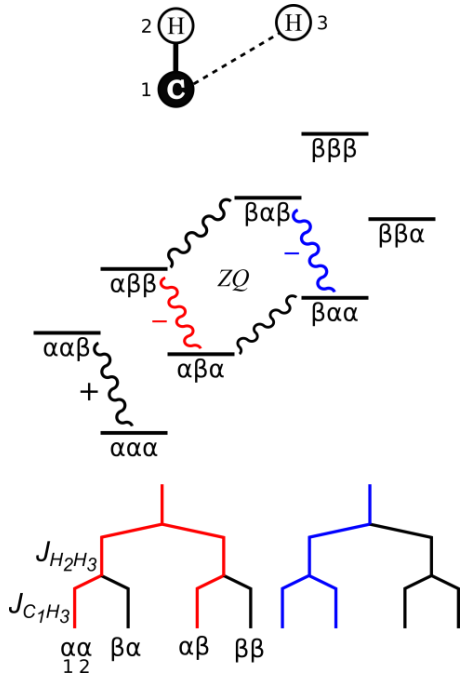



**Figure 4.7. Representation of the change in polarization between the two S$^3$ pulse sequences.**

**Top: Relevant (numbered) spins.**

**Middle: The energy levels (not to scale) of the 3 spins, the observed transitions in the homo (red) and hetero (blue) experiments (black transition is shared and used in both experiments) and the effect of the zero quantum (*ZQ*) pulse sequence. The associated linear combination is indicated by + and -.**

**Bottom: The active couplings in the homo (red) and hetero (blue) experiment are indicated in a coupling tree for spin 3 for a case where $J_{HH} > J_{CH}$.**

The new pulse train was established from the needed zero-quantum coherence polarization transfer. Theoretically there are two zero-quantum operators, $2\hat{I}_y\hat{S}_y+2\hat{I}_x\hat{S}_x$ and $2\hat{I}_y\hat{S}_x-2\hat{I}_x\hat{S}_y$, and the operator $2\hat{I}_y\hat{S}_y+2\hat{I}_x\hat{S}_x$ was chosen.[37,175] This needed to be translated into a pulse sequence, which was achieved by the following equations, starting from the $ZQ_x$ operator in equation (4.2) and (4.3). The transformations used are exemplified here, and explained further in Appendix A6.

$$ZQ_x = 2\hat{I}_y\hat{S}_y + 2\hat{I}_x\hat{S}_x \quad \textbf{(4.2)}$$

$$e^{-i\pi ZQ_x} = e^{-i\pi(2\hat{I}_y\hat{S}_y+2\hat{I}_x\hat{S}_x)} = e^{-i\pi(2\hat{I}_y\hat{S}_y)}e^{-i\pi(2\hat{I}_x\hat{S}_x)} \quad \textbf{(4.3)}$$

The operator $e^{-i\pi(2\hat{I}_y\hat{S}_y)}$ was extended to equation (4.4), as $e^k e^{-k} = 1$.

$$e^{-i\pi(2\hat{I}_y\hat{S}_y)} = e^{-i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}e^{i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}e^{-i\pi(2\hat{I}_y\hat{S}_y)}e^{-i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}e^{i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)} \quad \textbf{(4.4)}$$

The $e^{i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}e^{-i\pi(2\hat{I}_y\hat{S}_y)}e^{-i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}$ part was rewritten in equation (4.5).[175]

$$\begin{aligned} e^{i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}e^{-i\pi(2\hat{I}_y\hat{S}_y)}e^{-i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)} &= e^{\left[-i\pi e^{i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}(2\hat{I}_y\hat{S}_y)e^{-i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}\right]} \\ &= e^{\left[-i\pi e^{i\frac{\pi}{2}(\hat{I}_x)}e^{i\frac{\pi}{2}(\hat{S}_x)}(2\hat{I}_y\hat{S}_y)e^{-i\frac{\pi}{2}(\hat{I}_x)}e^{-i\frac{\pi}{2}(\hat{S}_x)}\right]} \end{aligned} \quad \textbf{(4.5)}$$

Using that $e^{i\varphi B}Ae^{-i\varphi B} = A\cos\varphi - i[A,B]\sin\varphi$, the $e^{i\varphi B}Ae^{-i\varphi B}$ $e^{i\frac{\pi}{2}(\hat{S}_x)}(2\hat{I}_y\hat{S}_y)e^{-i\frac{\pi}{2}(\hat{I}_x)}$ part of equation (4.5) was equal to equation (4.6).[175]

$$e^{i\frac{\pi}{2}(\hat{I}_x)}\left(2\hat{I}_y\hat{S}_y\right)e^{-i\frac{\pi}{2}(\hat{I}_x)} = -i\left[\hat{I}_x, 2\hat{I}_y\hat{S}_y\right] = -ii\left[2\hat{I}_z\hat{S}_y\right] = \left[2\hat{I}_z\hat{S}_y\right] \quad \textbf{(4.6)}$$

Similarly,

$$e^{i\frac{\pi}{2}(\hat{S}_x)}\left(2\hat{I}_z\hat{S}_y\right)e^{-i\frac{\pi}{2}(\hat{S}_x)} = -i\left[\hat{S}_x, 2\hat{I}_z\hat{S}_y\right] = -ii\left[2\hat{I}_z\hat{S}_z\right] = \left[2\hat{I}_z\hat{S}_z\right] \quad \textbf{(4.7)}$$

These answers were returned to equation (4.5), which was now equal to equation (4.8).

$$e^{\left[-i\pi e^{i\frac{\pi}{2}(\hat{I}_x)}e^{i\frac{\pi}{2}(\hat{S}_x)}(2\hat{I}_y\hat{S}_y)e^{-i\frac{\pi}{2}(\hat{I}_x)}e^{-i\frac{\pi}{2}(\hat{S}_x)}\right]} = e^{-i\pi(2\hat{I}_z\hat{S}_z)} \quad \textbf{(4.8)}$$

Which led to the first operator in equation (4.3) being equal to equation (4.9).

$$e^{-i\pi(2\hat{I}_y\hat{S}_y)} = e^{-i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}e^{-i\pi(2\hat{I}_z\hat{S}_z)}e^{i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)} \quad \textbf{(4.9)}$$

A similar approach was used throughout, done in Appendix A6, which led to equation (4.10).

$$\begin{aligned} e^{-i\pi(2\hat{I}_y\hat{S}_y+2\hat{I}_x\hat{S}_x)} &= e^{-i\pi(2\hat{I}_y\hat{S}_y)}e^{-i\pi(2\hat{I}_x\hat{S}_x)} \\ &= e^{-i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}e^{-i\pi(2\hat{I}_z\hat{S}_z)}e^{i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}e^{-i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)}e^{-i\pi(2\hat{I}_z\hat{S}_z)}e^{i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)} \\ &= e^{-i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}e^{-i\pi(2\hat{I}_z\hat{S}_z)}e^{-i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)}e^{i\frac{\pi}{2}(\hat{I}_z+\hat{S}_z)}e^{-i\pi(2\hat{I}_z\hat{S}_z)}e^{i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)} \\ &= e^{-i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}e^{-i\pi(2\hat{I}_z\hat{S}_z)}e^{-i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)}e^{-i\pi(2\hat{I}_z\hat{S}_z)}e^{i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)} \end{aligned} \quad \textbf{(4.10)}$$

Finally, it was used that –x = 270° x = 90° x + 180° x etc. to convert the operators into the pulse train seen in Figure 4.8. The optimal delays between the new pulses are $\tau/2$ or $(4\ {}^1J_{CH})^{-1}$, in accordance with polarization transfer of methine groups.[37]

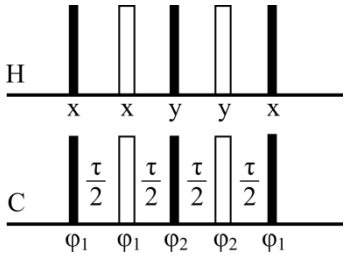



**Figure 4.8. The new zero quantum coherence pulse sequence element, which converts the polarization of the S$^3$ HMBC *homo* to achieve the *hetero* experiment. $\varphi_1$ = {x, -x, -x, x}, $\varphi$2 = {y, -y, -y, y}.**

### 4.3.2 Pulse sequences

The initial proposal for the S$^3$ HMBC *hetero* was to substitute the final $^{13}$C $\pi$/2 pulse in the original S$^3$ HMBC *homo* experiment with a new pulse train, as seen in Figure 4.9. Note that the first $\pi$/2 pulse is omitted in the carbon channel, since the carbon spins need to be aligned along *z* before detection.

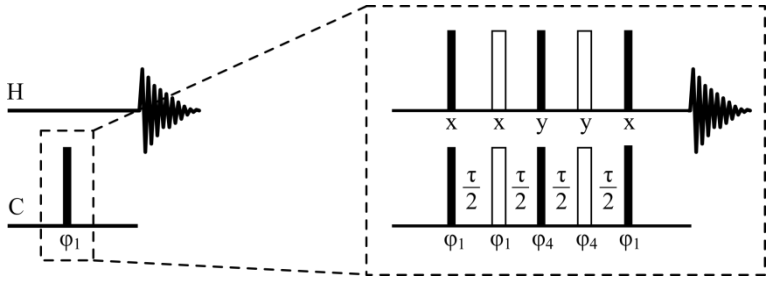



**Figure 4.9. The extended pulse sequence which replaces the final $\pi$/2 $S_x$ pulse of the S$^3$ HMBC to get the *hetero* experiment. $\varphi_1$ = {x, -x, -x, x}, $\varphi_4$ = {y, -y, -y, y}.**

In addition to inserting the new pulse sequence, the gradients of the S$^3$ HMBC *homo* experiment (a) in Figure 4.5, had to be adjusted, by extension to select $G_1(H^{-1}/C^{+1})$-$G_2(H^{-1}/C^{-1})$-$G_3(H^{-1}/C^{0})$ instead of $G_1(H^{-1}/C^{+1})$-$G_2(H^{-1}/C^{-1})$. This was achieved by using a gradient power ratio of 2:-2:1 instead of 5:-3 for echo, and -2:2:1 compared to -3:5 for anti-echo. Upon appending the new pulse train to obtain a new pulse sequence it was necessary to ensure that the carbon chemical shifts are refocused between the first $^{13}$C $\pi$/2 (disregarding the low pass *J*-filter), and the first $^{13}$C $\pi$/2 pulse in the new pulse train. This was in principle achieved in two ways which should both lead to refocusing, as illustrated in Figure 4.10, using sequence (a) as an example. In short, the first version withdraws the time of a $^1$H $\pi$/2 pulse from the delay after the first $^{13}$C $\pi$ pulse to ensure refocusing of $\delta_C$, while the $\pi$/2 pulses in the pulse train are offset in the second version in a manner that ensures refocusing.

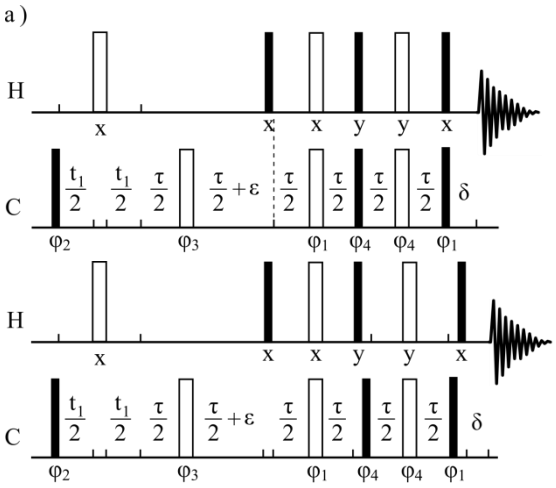



**Figure 4.10. Representation of two possible methods to obtained needed adjustments of the delays for carbon chemical shift refocusing, exemplified for sequence (a).**

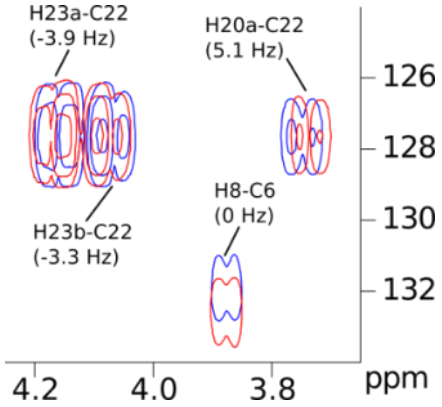



**Figure 4.11. Example of methine and quaternary carbon displacement in the $S^3$ HMBC *hetero* methine/methyl spectrum of strychnine (180 mM).**

General rules when combining data for pulse sequences needed to be obeyed as well:

1. The $^{13}C$ chemical shift was refocused for the minimum $t_1$.
2. Sample heating due to decoupling was equal.
3. The evolution times for $\delta_H$, $J_{HH}$ and $J_{CH}$ were equal.
4. The lengths of the sequences were identical.

All of these requirements were met in the $S^3$ HMBC, and since an identical pulse train was appended, all should be met by design.

Unfortunately this strategy was not successful. While the polarization transfer worked and led to extraction of $^nJ_{CH}$-coupling constants, the refocusing of carbon chemical shifts was imperfect. The refocusing problems for methine groups were negligible and resulted only in slight broadening of the resonances in the F1 dimension in the add/subtract spectra. For quaternary carbons the problem was more pronounced, which is exemplified in Figure 4.11, and even though these are of no interest in this experiment, the problem needed to be solved, to ensure the extraction of correct coupling constants throughout. Note also that the elimination of quaternary carbons in the methine/methyl spectra by the multiplicity editing did not work properly as a consequence.

The next approach taken was to append the new pulse train after the entirety of the original homonuclear $S^3$ HMBC, as the refocusing here worked perfectly, leading to inherent refocusing of the new pulse sequence. The gradients were rearranged to achieve an assumed clean coherence selection as seen in Figure 4.12, top left. This adjustment did not initially give the desired refocusing of carbon chemical shifts though. The gradient selection was identified to be the cause of the problem, as the top right sequence in Figure 4.12 led to messy but refocused spectra. Thus, a new gradient scheme was developed in order to alleviate this problem, as illustrated in Figure 4.12 bottom left for sequence (a), while similar corrections were made for sequence (b-e). Note that a similar gradient scheme was not possible for the first proposed pulse sequence, due to a mismatch of gradient duration and delays in some of the sequences.

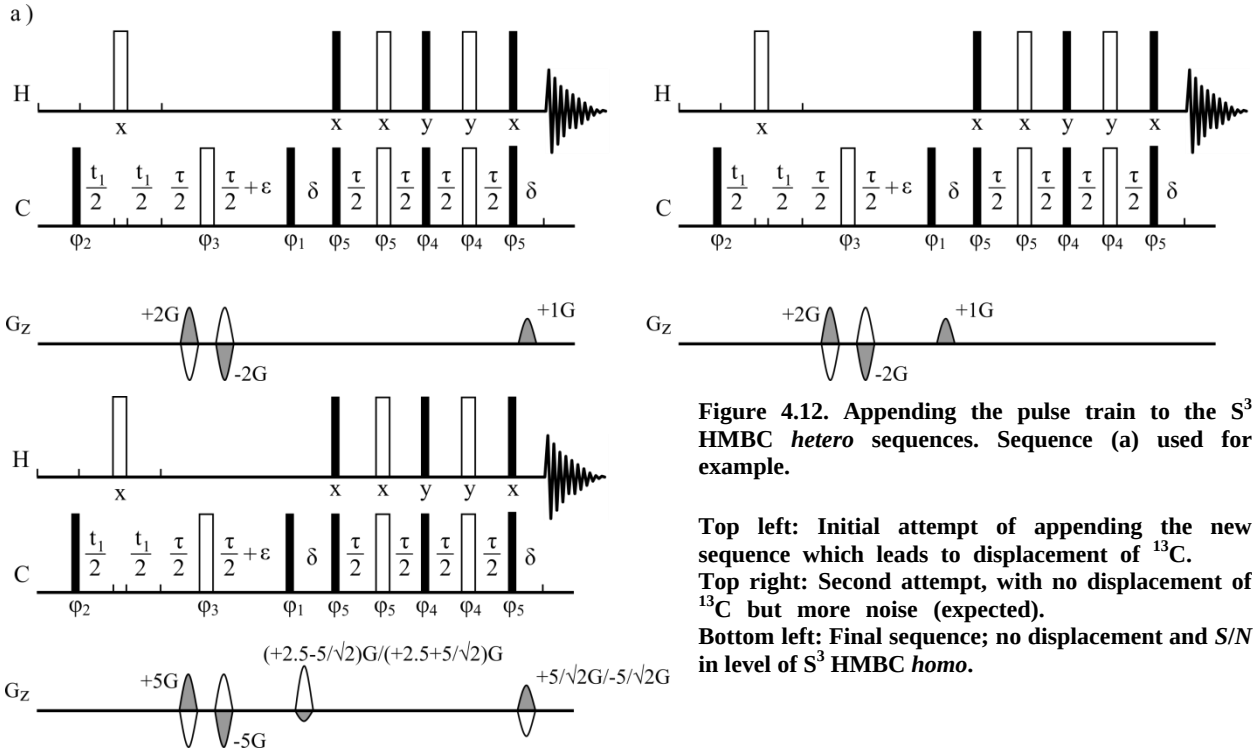



**Figure 4.12. Appending the pulse train to the S³ HMBC *hetero* sequences. Sequence (a) used for example.**

**Top left: Initial attempt of appending the new sequence which leads to displacement of $^{13}C$.**
**Top right: Second attempt, with no displacement of $^{13}C$ but more noise (expected).**
**Bottom left: Final sequence; no displacement and *S/N* in level of S³ HMBC *homo*.**

It is not clear why the new gradient scheme is superior, as the chosen coherence pathway is supposedly identical between all sequences. The coherence selection by the gradients in the two first sequences in Figure 4.12 is $G_1(H^{-1}/C^{+1})$-$G_2(H^{-1}/C^{-1})$-$G_3(H^{-1}/C^{0})$. In the final sequence an extra gradient pair was added to choose coherence $G_1(H^{-1}/C^{+1})$-$G_2(H^{-1}/C^{-1})$-$G_3(H^{-1}/C^{0})$-$G_4(H^{-1}/C^{0})$, or in words, make sure that the change in coherence order was zero for both protons and carbons across the new pulse train. Experimental evidence showed that it was needed to gain this further control of the coherence selection, which in practice was achieved by using gradients as in the S³ HMBC, only changing sequence (a) as earlier described, and inserting new gradients around the new pulses with opposite signs. The first new gradient is added to the last existing gradient from the S³ HMBC *homo* sequence, as indicated in Figure 4.12.

Appending the new pulse train led to an increase in the phase cycle compared to just exchanging the last $^{13}C$ $\pi/2$ pulse, as the phases of the new element were cycled independently of $\varphi_1$. This was not strictly needed as the spectra and coupling constants using $\varphi_1 = \varphi_5$ and $\varphi_1 \neq \varphi_5$ were essentially identical. It is thus recommended that for four scan (ns) experiments $\varphi_1$ is set equal to $\varphi_5$, while $\varphi_1$ should be cycled independently of $\varphi_5$ for ns > 4 scans. In both cases $\varphi_2$ and $\varphi_3$ were cycled independently and should be changed accordingly. Since the pulse train was appended and identical for all pulse sequences, previously mentioned requirements for combining pulse sequences were inherently met.

The new pulse sequences are found in Figure 4.16 and the result of the change in polarization is exemplified in Figure 4.13 by the simple compound vinyl acetate. The coupling constants are compared to 1D and reference values to an excellent correlation in Figure 4.14 and Table 4.3.[176]

**Table 4.3. Comparison of $S^3$ HMBC coupling constants of vinyl acetate to 1D $^1$H spectra (*a*) and literature (*b*).[176]**

| | $^3J_{HH}$ [a] [Hz] | | $^2J_{CH}$ [b] [Hz] | |
|---|---|---|---|---|
| | $S^3$ HMBC *homo* | Ref. | $S^3$ HMBC *hetero* | Ref. |
| **H-4a** | +6.3 | 6.2 | +7.4 | +7.6 |
| **H-4b** | +14.0 | 14.0 | -7.9 | -7.9 |

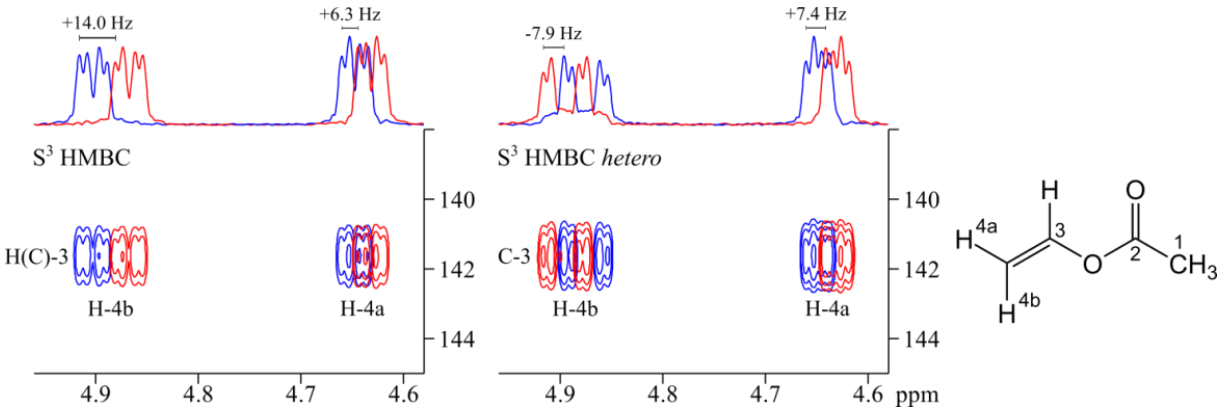



**Figure 4.13. Left: Comparison of $S^3$ HMBC *homo* and $S^3$ HMBC *hetero* spectra for C(H)-3 of vinyl acetate (220 mM).**

To better understand the coupling pattern of the homo- and heteronuclear variants of the $S^3$ HMBC the cross-peaks of C-3/H-4b and C-3/H-4a are explained in Figure 4.14. As seen both $J_{CH}$ and $J_{HH}$ were part of the multiplet in both experiments and if the (+) and (–) spectra were overlaid, the combined shape of the multiplets was almost identical.

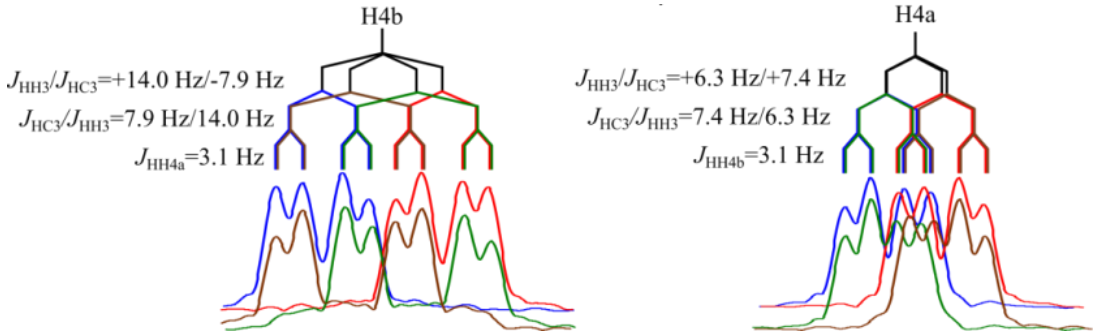



**Figure 4.14. Comparison of $S^3$ HMBC *homo* (blue/red) and $S^3$ HMBC *hetero* (green/brown) multiplet for H-4b and H-4a of vinyl acetate. Black lines in coupling tree indicate $S^3$ editing. The coupling trees are offset slightly horizontally to better distinguish colors.**

The in-phase $J_{H4bH4a}$ coupling constant was determined to be larger for both experiments than the 1.5 Hz determined from a $^1$H spectrum, possibly due to a lack of resolution of the $S^3$ HMBCs with 2.5 Hz/point (4k versus 64k points acquired for 1D $^1$H spectrum). As expected the coupling trees were alike with the difference being the coupling constant leading to $S^3$ editing, where sign determination was possible.

*J*-crosstalk was experimentally more apparent, for a few resonances, compared to the native $S^3$ HMBC where *J*-crosstalk was always in the noise level. The *J*-crosstalk may alter the experimentally determined coupling constants and will always lead to the extraction of smaller coupling constants and needed to be addressed. As for the homonuclear experiment, the *J*-crosstalk

could be minimized by scaling [b+c] with a factor *k* and [d+e] with $(2k/[k+1])^{½}$. If scaling was not sufficient the crosstalk was diminished by scaled combinations of the α and β spectra, as described earlier for $S^3$ experiments.[165] The subspectra of vinyl acetate were the only spectra where *J*-crosstalk was immediately apparent, even when subspectra of other compounds were scrutinized. In Figure 4.15 the standard editing of the spectra is compared to spectra edited by a scaled combinations of the α and β spectra to decrease *J*-crosstalk.[165] Most importantly the extracted *J*-coupling constants varied by less than 0.1 Hz, and thus the differences were lower than the expected experimental uncertainty of these kinds of experiments. It is acknowledged that the difference may be larger for different compounds as described in the literature.[177]

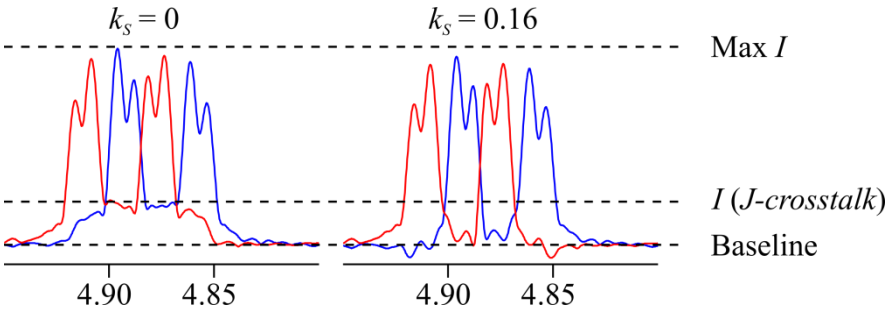
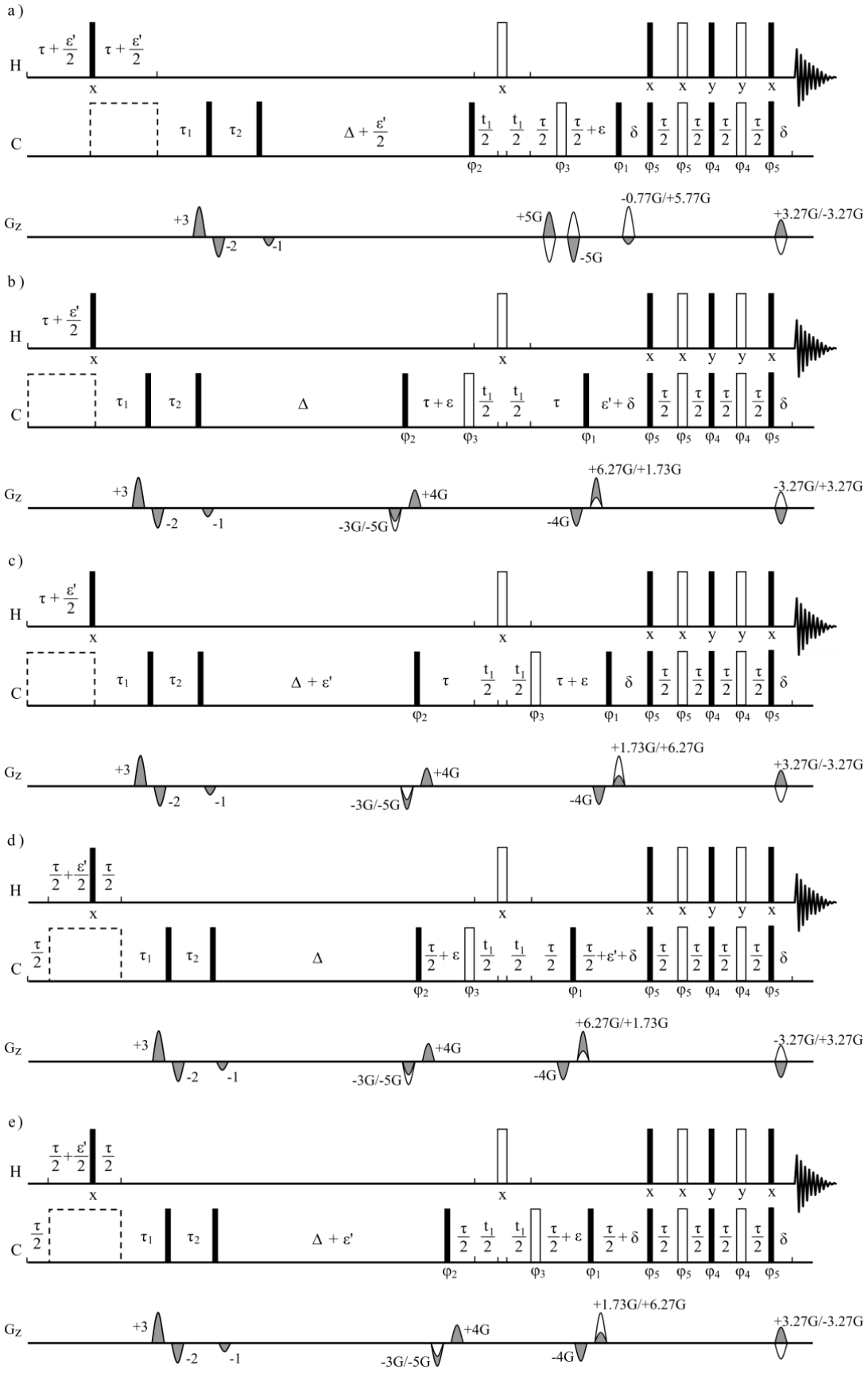



**Figure 4.15. Comparison of the *J*-crosstalk in standard and edited spectra of the C3-H4a cross-peak of vinyl acetate (220 mM). Edited by $\alpha_S = \alpha - k_S \cdot \beta$ and $\beta_S = \beta - k_S \cdot \alpha$. The difference in the extracted $J_{CH}$-coupling constant is below 0.1 Hz.**

**Figure 4.16. Next page. $S^3$ HMBC *hetero* pulse sequence comprised by modified sequences a-c; edited HMBC sequences, and d,e; HAT HMBC sequences, all shown with a 2nd-order LPJF, which may be exchanged to 3rd-order LPJF or removed.**

**Filled and open bars refer to π/2 and π pulses, respectively, and the dashed open boxes represent $^{13}C$ decoupling. $\tau = (2\ ^1J_{CH})^{-1}$ or $(^1J_{max} + {}^1J_{min})^{-1}$, δ = gradient delay, $\varepsilon = 2\ t_{1/2,min} + t(\pi_H)$, $\varepsilon' = \varepsilon + t(\pi_C)$, $\tau_1 = ½[^1J_{min} + 0.146\ (^1J_{max} - {}^1J_{min})]^{-1}$, $\tau_2 = ½[^1J_{max} - 0.146\ (^1J_{max} - {}^1J_{min})]^{-1}$, Δ = delay for heteronuclear long-range coupling evolution. $(2 \cdot 8\ Hz)^{-1}$ is standard.**

**Phase cycles. $\varphi_1$ = {x, x, 4(-x), x, x}, $\varphi_2$ = {4(x), 8(-x), 4(x)}, $\varphi_3$ = {8(x), 8(y), 8(-x), 8(-y)}, $\varphi_4$ = {x, -x, -x, x}, $\varphi_5$ = {y, -y, -y, y}**


a)
H
C
Gz
x
y
τ1
τ2
Δ + ε'/2
t1/2
τ/2
τ/2 + ε
δ
φ2
φ3
φ1
φ5
φ4
+3
-2
-1
+5G
-5G
-0.77G/+5.77G
+3.27G/-3.27G
b)
Δ
τ + ε
τ
ε' + δ
+6.27G/+1.73G
-3.27G/+3.27G
+4G
-3G/-5G
-4G
c)
Δ + ε'
+1.73G/+6.27G
d)
τ/2 + ε'/2
τ/2 + ε'+ δ
e)

### 4.3.3 Computation of long-range heteronuclear coupling constants

Unlike the homonuclear long-range coupling constants, the heteronuclear variants have been more extensively investigated.[46,55,135,145,146] Unfortunately, a study similar to that of Bally and Rablen was not found, and the reported calculations use a wide variety of methods and functionals, some of which seem to benefit from scaling.[53,55] This is in principle not a problem, but the lack of a true and tried methodology and scaling factor may lead to overfitting as the computed coupling constants are scaled to fit only the current data or a couple of datasets. However the method of "local scaling" is generally used in the calculation of chemical shifts, as discussed in Section 2.4.3.[47,54,178]

Investigating and correlating calculations of long-range heteronuclear coupling constants to a large set of experimental data was not a part of the work; in part due to discrepancies between methods to extract coupling constants. It should be noted though, that scaling is often required for accurate determination of NMR observable from DFT calculation, probably due to a constant and inherent error in the calculations. It is thus often advantageous to use the correlation factor to assess the fit between experimental and theoretical data, as the correlation factor is independent of needed scaling.

Two methods were investigated with different functionals; B3LYP, which worked nicely for homonuclear coupling constants, and MPW1PW91, which is very often used specifically to better calculate NMR properties.[46,47,53,54,179,180] For both methods it was apparent that a larger basis set improved the correlation between experimental and calculated data, while e.g. diffuse functions had little to no effect. It was necessary to use the full theoretical *J*-coupling constants and not just the Fermi contact (FC) to get the best results, while the FC term was still dominating in the calculations. This was in contrast to the method used for calculations of $J_{HH}$ by DFT. For more on different components used in *J*-coupling constants calculations by DFT see Section 2.4.4.

MPW1PW91 generally resulted in a better correlation to the experimental data when used to calculate coupling constants while the difference between optimized structures from B3LYP and MPW1PW91 differed little, as long as identical basis sets were used. In the end MPW1PW91/6-311G(d,p) was used for both optimization and *J*-coupling constant calculations as this resulted in the better reproduction of the experimental *J*-coupling constants. Comparisons may be found in Appendix A6.

### 4.3.4 Results

The $S^3$ HMBC *hetero* experiment was as the homonuclear variant tested on strychnine. The coupling constants were easily extracted from displacements of 1D slices from cross-peaks equivalent to the homonuclear $S^3$ HMBC. Examples of cross-peaks as well as 1D slices are found in Figure 4.17 and a full spectrum in Appendix A6.

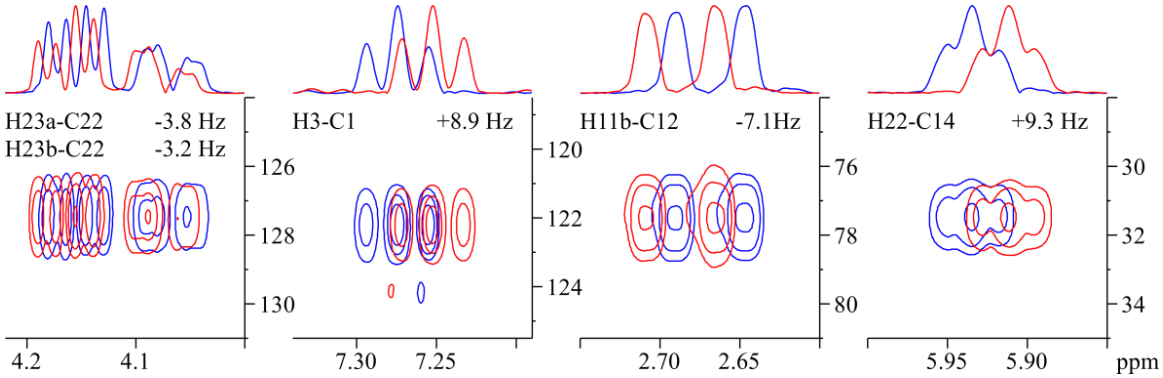



**Figure 4.17. Excerpts and cross-sections through F1 of cross-peaks of S3 HMBC *hetero* for strychnine (180 mM). Note that the sign of the coupling constants are apparent from the relative position of the cross-peaks in the subspectra.**

The extracted coupling constants correlate well to theoretical coupling constants, especially if the calculations are scaled, and a comparison to theoretical as well as literature values is presented in Table 4.4 and Figure 4.18.

**Table 4.4. Experimental $S^3$ HMBC *hetero*, literature and theoretical $^nJ_{CH}$-coupling constants for strychnine.[153,160,181,182] All *J*-coupling constants are in Hz. Both optimization and *J*-coupling constant calculations were performed at a MPW1PW91/6-311+g(d,p) level of theory. Calculated coupling constants are given without scaling and linearly scaled to the experimental data as $J_{scaled} = (J_{calc}-b)/a$. $a = 0.92$, $b = -0.14$.**

| H | C | Experimental (Hz) | | | | | Theoretical (Hz) | | n |
|---|---|---|---|---|---|---|---|---|---|
| # | # | S3 HMBC *hetero* | [160] | [181] | [153] | [182] | Calc. | Scaled | |
| 1 | 3 | 7.4 | 7.5,7.2 | | | | 6.7 | 7.4 | 3 |
| 2 | 4 | 7.9 | | | | | 7.0 | 7.8 | 3 |
| 3 | 1 | 8.9 | 8.9,9.3 | | | | 8.1 | 8.9 | 3 |
| 4 | 2 | 7.7 | | 7.5,7.4 | 7.9 | | 7.1 | 7.8 | 3 |
| 8 | 12 | 6.4 | 5.8,5.6 | 5.5,5.6 | 5.3 | | 5.6 | 6.2 | 3 |
| 8 | 13 | -1.7 | | | | | -1.7 | -1.6 | 2 |
| 11b | 12 | -7.1 | 6.9,-6.9 | 7,6.9 | 6.8 | | -6.9 | -7.3 | 2 |
| 13 | 8 | -6.3 | | 6.3,6.4,6.3 | 6.2 | | -6.0 | -6.3 | 2 |
| 13 | 14 | -4.9 | 5.5,-4.5 | 4.6,4.7 | 4.5 | 4.7 | -4.7 | -5.0 | 2 |
| 15a | 13 | 7.7 | | 8,8.1 | 8.0 | 7.9 | 7.0 | 7.7 | 3 |
| 15a | 14 | -2.9 | 1.2,-2.9 | 1.8,2.8,2.8 | 2.0 | 3.2 | -2.5 | -2.6 | 2 |
| 15a | 16 | -4.6 | | 4.5,4.3 | 4.0 | | -4.3 | -4.5 | 2 |
| 15b | 14 | -3.2 | 2.3,-2.6 | 3.2,3.6 | 2.8 | 4.8 | -3.1 | -3.2 | 2 |
| 15b | 16 | -1.9 | | 2.7,3 | 1.4 | | -2.1 | -2.2 | 2 |
| 16 | 14 | 6.2 | 6.7,6.4 | | | 6.4 | 5.8 | 6.4 | 3 |
| 20a | 14 | 1.4 | | | | | 1.3 | 1.6 | 3 |
| 20a | 22 | 5.6 | 4.7,5 | 6.1,5.8 | 5.1 | | 4.6 | 5.1 | 3 |
| 20b | 14 | 5.5 | 5.7,5.4 | 5.4,5.3 | 5.1 | 5.5 | 4.7 | 5.2 | 3 |
| 20b | 16 | 7.8 | | 6.9,6.9 | 6.7 | | 6.3 | 6.9 | 3 |
| 20b | 22 | 4.6 | 5.1,5.5 | 4.5,4.6 | 4.2 | | 4.6 | 5.1 | 3 |
| 22 | 14 | 8.5 | 8.5,8.6 | 7.9,8.5,8.6 | 8.8 | 8.9 | 8.0 | 8.8 | 3 |
| 23a | 22 | -3.8 | 4.2,-3.8 | 3.8,4 | | | -3.8 | -3.9 | 2 |
| 23b | 22 | -3.4 | 3.6,-3.4 | 4,3.9 | | | -3.0 | -3.1 | 2 |
| | | | | | RMSD | | 0.53 | 0.22 | |
| | | | | | $R^2$ | | 0.999 | | |

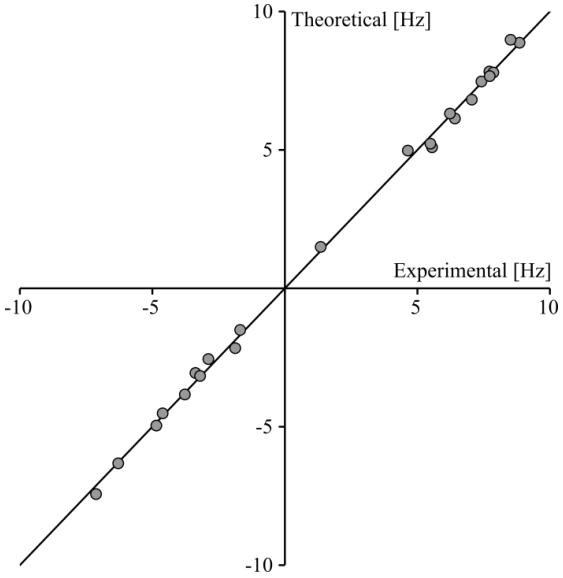



**Figure 4.18. Experimental versus DFT-calculated coupling constants, where the line represents a perfect fit. Calculated coupling constants are linearly scaled to the experimental data as $J_{scaled} = (J_{calc}-b)/a$. $a = 0.92$, $b = -0.14$.**

The extracted coupling constants were in excellent agreement with previously published data and scaled theoretical calculations, with an RMSD of 0.22 Hz when comparing to the latter. This is almost equal to the RMSD determined for homonuclear coupling constants from $S^3$ HMBC (0.21 Hz), indicating the same general precision of the two methods. When comparing to the theoretical calculations, note also the high correlation factor. This is in support of the differences between experimental and calculated coupling constants being due to an inherent and consequent error in the DFT calculations, which may be alleviated by scaling consistent with literature findings.[53] A conclusion, which was supported by differences in scaling factors *a* and *b* between functionals, while a high $R^2$ was generally retained. It could of course also be the product of a constantly erroneous extraction of coupling constants but good correlation to literature values contradicts this.[153,160,181,182] It would have been beneficial to have a golden standard for comparison as to determine whether the deviation was a result of the experiment or the theoretical method utilized. Like the homonuclear $S^3$ HMBC it was possible to extract coupling constants across heteroatoms, which is important in e.g. carbohydrate 3D structural investigations.[15,135,183] This is a nice feature of HMBC compared to TOCSY type experiments, as the latter utilizes $^3J_{HH}$ to generate the long-range coupling constants.[145]

The $S^3$ HMBC *hetero* experiment, and in large regard also the homonuclear $S^3$ HMBC experiment, has one inherent flaw; it is only really a viable option for coupling constants involving methine carbons. While this may limit widespread usage, the high accuracy, ease of extraction and sign-selective nature of the experiment should cater for a lot of fields where the resulting information will immediately increase the structural knowledge. An obvious field is that of carbohydrate chemistry, where coupling constants across the glycosidic linkage may be used for conformer analysis.[119,135,183,184]

## 4.4 Perspectives

As interesting as the possibility to obtain the long-range coupling constants was, the greater question is whether the observed long-range coupling constants could actually discern stereoisomers. The determined long-range $J_{HH}$ values were small and the information may be non-discerning and thus of lesser value. To assess this, a probability function was considered attractive.

### 4.4.1 Establishing a probability function for *J*

The DP4 is in the opinion of the author, a very informative way to evaluate diastereomers from chemical shift differences, see Section 3.1.6.[58] A similar probability function, here dubbed JP4, would be desirable for *J*-coupling constants especially as DFT calculations become easier available.

A probability function was thus established from coupling constants obtained from the literature, as determining these from scratch was not in the scope of the project and was deemed too time consuming. The dataset published by Bally and Rablen was used as the combined dataset of test and the two probe sets (combined to 228 coupling constants from multiple structures).[53] The standard deviation from the *J*-coupling constants of their dataset is 0.51 Hz.[53] The intercept of the dataset was 0.0026 Hz, and is approximated to 0 Hz in the following. The degrees of freedom ($\nu$) parameter was harder to determine and it was only estimated in the current work.[III] The data set was fitted to a standard student t-test curve to a value where the difference in the fit of $\nu$ and $\nu$-1 was below 0.01. This equals a value of 8.76 (~9). It should be noted that $\nu$ does not drastically influence the result of the evaluation, except when a very low or extremely high value is chosen. The formula for the resulting "JP4" probability function is given in (4.11).

$$P(i|J_N) = \frac{\prod_{k=1}^{N} 1 - T^{\nu}\left(\left|J_{scaled,k}^{-i} - J_{exp,k} - \mu\right|/\sigma\right)}{\sum_{j=1}^{m}\prod_{k=1}^{N} 1 - T^{\nu}\left(\left|J_{scaled,k}^{-i} - J_{exp,k} - \mu\right|/\sigma\right)} \qquad \textbf{(4.11)}$$

With $\mu$ = 0, $\nu$ = 9 and $\sigma$ = 0.51 Hz. The number of data points compared to the DP4 probability function was rather sparse with 228 *J*-coupling constants compared to 1717 $^{13}$C and 1794 $^{1}$H, and thus the resulting probability will be prone to more errors. The JP4 value will thus never be used alone, but it is a very intuitive way to compare experimental and theoretical data sets. It should also be noted that this is of course only valid if the methodology of Bally and Rablen is used.[53] The output of the JP4 function is the probability in percent that a given stereoisomer is correct according to the data.

### 4.4.2 Differentiation of stereoisomers by S$^3$ HMBC

To check whether the long-range coupling constants were discerning of stereoisomers the data was purposely calculated for wrong assignment of diastereotopic protons for strychnine.

Starting from the homonuclear coupling constants the correlation between the full and a reduced set of experimental coupling constants and theoretical values are seen in Table 4.5. The reduced set is

[III] $\nu$ was determined from a statistics program based on the data set in the original DP4 publication, which was not available to the author. $\nu(^{1}H) = 11.38$, $\nu(^{13}C) = 14.18$.[58]

the data with all $^3J_{HH}$-coupling constants removed, as these could on the most part be extracted with other methods.

**Table 4.5. Comparison of correct and incorrect assignment of the diastereotopic protons in strychnine. The full set includes all coupling constants extracted by S$^3$ HMBC *homo*, while $^3J_{HH}$-coupling constants have been removed in the reduced set. The number on the left indicates the diastereotopic protons which are switched compared to the correct assignment.**

| | **Full** | | **Reduced** | |
|---|---|---|---|---|
| Switched # | MAE | JP4 | MAE | JP4 |
| Right | 0.15 | 99.9 | 0.06 | 47.9 |
| Dia-11 | 0.38 | 0.0 | 0.08 | 0.9 |
| Dia-15 | 0.38 | 0.0 | 0.09 | 0.0 |
| Dia-18 | 0.17 | 0.1 | 0.09 | 0.0 |
| Dia-20 | 0.23 | 0.0 | 0.15 | 0.0 |
| Dia-23 | 0.18 | 0.0 | 0.06 | 51.2 |
| Dia-All | 0.75 | 0.0 | 0.22 | 0.0 |

The mean average error (MAE) was not necessarily a good method to differentiate the datasets as it was hard to evaluate whether the differences were within expected errors. Here the JP4 function made it much easier to quickly evaluate the data. Even without the $^3J_{HH}$-coupling constants it was possible to assign all diastereotopic proton pairs using the JP4 probability function except for two: 17, where the protons overlap, and 23, where the two observed coupling constants available were almost equal (both expected to be positive and under 0.1 Hz). It should be noted that, as it is a probability function, the values of the right assignment and the misassignment of 23 meant that they could not be discerned. Still, it was a testimony to the high degree of accuracy and the importance of identifying the sign in long-range coupling constant experiment that the remaining diastereomers could be differentiated, underlining the vast possibilities for using this experiment for 3D structural elucidation. All of the coupling constants in the reduced set were thus between -3.0 and 1.1 Hz, and very exact coupling constants were needed to facilitate differentiation.

In a similar fashion the diastereotopic protons of strychnine were again switched for the heteronuclear data, and the experiment was clearly discriminating, as seen in Table 4.6.

**Table 4.6. Comparison of right and wrong assignment of the diastereotopic protons in strychnine using $^nJ_{CH}$-coupling constants from the S$^3$ HMBC *hetero* experiment.**

| **#** | **MAE** | **JP4** |
|---|---|---|
| - | 0.24 | 94.1 |
| 11 | 0.88 | 0.0 |
| 15 | 0.66 | 0.0 |
| 20 | 0.75 | 0.0 |
| 18 | 1.02 | 0.0 |
| 23 | 0.29 | 5.9 |
| All | 2.63 | 0.0 |

It is recognized that using the JP4 probability function for comparison is clearly problematic; and that the value holds little meaning. But, as the standard deviation from the strychnine data was smaller than for the dataset used in the setup of the JP4 probability (and equal to coupling constants from the homonuclear S$^3$ HMBC variant),[53] and if a comparable degrees of freedom value is assumed, it is indicative of the discriminative prowess of the long-range $J_{CH}$-coupling constants.

The $^{n}J_{CH}$- and $^{n+1}J_{HH}$-coupling constants were clearly discriminative of diastereomers and the two experiments are believed to greatly increase the amount of available structural data and resulting structural knowledge.

### 4.4.3 $^{n}J_{CqH}$-coupling constants from S$^3$ HMBC *homo*

The major drawback of the S$^3$ HMBC *hetero* experiment was the inability to measure coupling constants to methylene or quaternary carbons. Especially the latter is problematic, as the structural knowledge gained from $^{n}J_{CH}$-coupling constants is increasingly needed as the proton-to-heteroatom ratio decreases. In that regard it is appreciable that $J_{CqH}$ coupling constants are implicitly embedded in the add spectrum of methylene and quaternary carbons in both S$^3$ HMBC experiments, without sign due to the lack of S$^3$ editing.[IV] The multiplet became complicated in cases where the proton couples to other protons, but in proton deficient compounds where the proton multiplets should be simple, maybe even singlets, the size of the coupling constants may be easily extracted and used in structure elucidation. An example of a rather complicated situation is given in Figure 4.19 where 1D slices from S$^3$ HMBC *homo* of strychnine are compared to theoretically simulated spectra using the given coupling constants. Theoretical heteronuclear coupling constants are, in order from left: 1.2, -5.1 and -7.3 Hz, calculated at a MPW1PW91/6-311+g(d,p) level of theory and scaled according to the methine data. The extracted coupling constants could thus be used in a large-medium-small analysis as suggested by Bifulco *et al* if not used directly.[46] In-phase homonuclear coupling constants were initially those from Table 4.2 and varied slightly to increase the fit between the experimental and theoretical spectra. The $^{2}J_{HH}$-coupling constant was estimated from 1D $^{1}$H spectra.

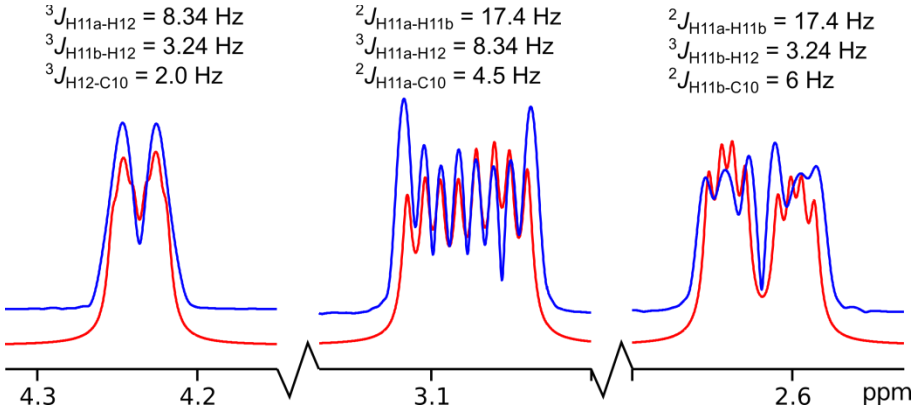



**Figure 4.19. 1D slices of couplings to C-10 in strychnine (300 mM) from S$^3$ HMBC *homo*. Experimental data (blue) compared to simulated spectra (red). Theoretical heteronuclear coupling constants are (from left): 1.2, -5.1 and -7.3 Hz. Spectral resolution in the S$^3$ HMBC *homo* was 2.5 Hz/point. Simulated using the coupling constants given above the resonances with a line broadening of 2.5 Hz using the Daisy utility in Topspin.**

[IV] Quaternary carbons cannot have their polarization interchanged from $^{1}$H to $^{13}$C, for obvious reasons.

### 4.4.4 Extraction of coupling constants to methylene (C)H

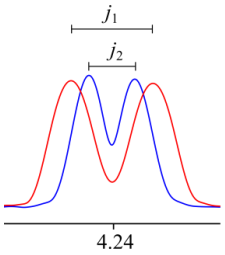
Unlike the easy extraction of coupling constants to methine groups, extraction of constants to methylene and methyl groups is more problematic, making this experiment a poorer choice for extractions of this kind of data. It seldom posed a significant problem in practice, since coupling constants that involve a methine and e.g. a methylene group in most cases could be extracted easily from the methine carbon. A simple approach to extract approximate coupling constants to methylene protons was identified, and given in equation (4.12), from measured distances between in-phase peaks ($j_1$ and $j_2$), illustrated in Table 4.7. It is especially easy when the passive coupling constants were small.

$$ {}^{n}J^{1}_{HH} = \frac{j_1 + j_2}{2}, \qquad {}^{n}J^{2}_{HH} = \left| {}^{n}J^{1}_{HH} - j_2 \right| \tag{4.12} $$

The equation yields the coupling constants as two active coupling constants lead to $S^3$ editing per coupling pattern, while the remaining coupling constants are in-phase and thus appear in both subspectra. The overlaid cross-peaks will thus have a pattern where the splitting in one sub-spectrum is equal to ${}^{n}J_{HH}{}^{1} + {}^{n}J_{HH}{}^{2}$ and the other $|{}^{n}J_{HH}{}^{1} - {}^{n}J_{HH}{}^{2}|$. Coupling constants extracted by this method are given in Table 4.7 and compared to methine data when possible.

**Table 4.7. Comparison of long-range coupling constants to methylenes in $S^3$ HMBC *homo*. Compared to the values extracted to methines from Table 4.2. *Theoretical coupling constants calculated as in Table 4.2. Right: The extraction of $j_1$ and $j_2$ from the 1D slices of cross-peaks.**

| (C)H1 | H2 | $j_1$ [Hz] | $j_2$ [Hz] | $J_1$ [Hz] | $J_2$ [Hz] | $J_{methine}$ [Hz] | Other [Hz] |
|---|---|---|---|---|---|---|---|
| 4a/b | 3 | 20.3 | 7.5 | 13.9 | 6.4 | 14.0,6.3 | ${}^{2}J_{CH}$=9.4 |
| 11a/b | 12 | 12.2 | 6.8 | 9.5 | 2.7 | 8.5, 3.4 | ${}^{2}J_{CH}$=1.4, ${}^{3}J_{HH}$=2.5 |
| 15a/b | 16 | 5.7 | 3.3 | 4.5 | 1.2 | 4.0,2.2 | ${}^{2}J_{CH}$=1.3 |
| 23a/b | 22 | 13.2 | 0.4 | 6.8 | 6.4 | 7.0,6.1 | ${}^{2}J_{CH}$=6.8, ${}^{4}J_{HH}$=4.0 |
| 17a/b | 18a | 10.1 | 2.6 | 6.3 | 3.7 | 7.5,0.5* | ${}^{2}J_{CH}$=1.0, ${}^{2}J_{HH}$=8.8 |
| 17a/b | 18b | 19.1 | 1.3 | 10.2 | 8.9 | 12.1,6.3* | ${}^{2}J_{CH}$=3.1, ${}^{2}J_{HH}$=10.1 |
| 18a/b | 17a/b | 16.2 | 4.8 | 10.5 | 5.7 | - | - |




In Figure 4.20 examples of simulated spectra based in coupling constants to methylene protons are depicted, from the determined values found in Table 4.7. When the coupling constants were large and different they were easily extracted, as seen on the left for vinyl acetate. As soon as the coupling constants became smaller and/or more equal, the extractions were much more challenging (right). Especially when more passive coupling constants were added to the in-phase coupling patterns, complications ensued.

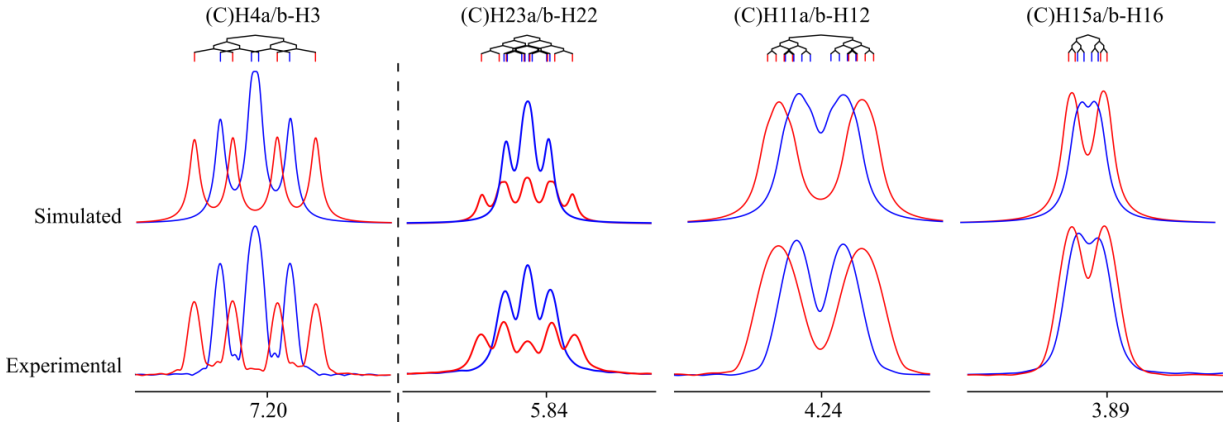



**Figure 4.20. Experimental and simulated multiplicity patterns for vinyl acetate (left, 220 mM) and strychnine (right, 300 mM). The simulated peaks are all obtained from a manual fitting in MestreNova. Coupling trees are indicative. Coupling constants found in Table 4.7.**

It was also possible to extract coupling constants between the 17 and 18 protons, as seen in Figure 4.21. The theoretically most complicated peak that includes 17a/b and 18a/b was seemingly the most simple as the $^{2}J$ methylene coupling constants were not observed and the coupling constants seemed to average.

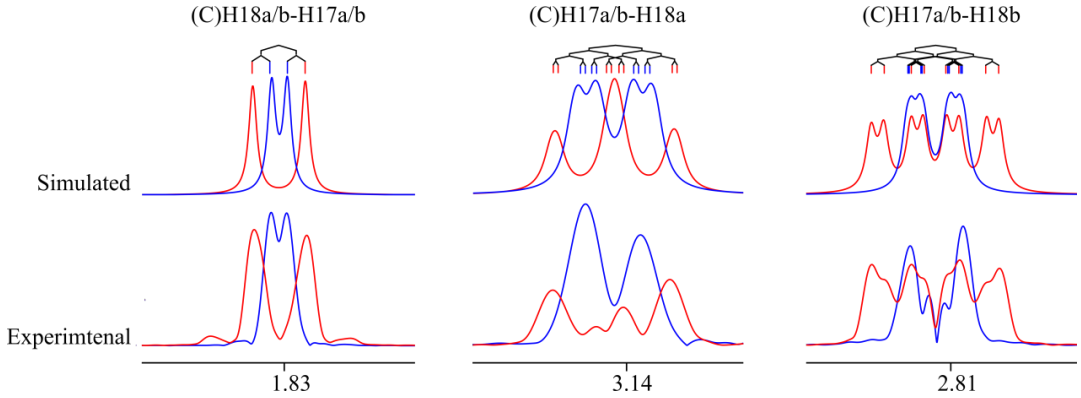



**Figure 4.21. Experimental and simulated multiplicity patterns from strychnine (300 mM). The simulated peaks are all obtained from a manual fitting of coupling constants in MestreNova. Coupling trees a indicative. Coupling constants found in Table 4.7.**

In conclusion, while it was possible to extract coupling constants to methylenes, the results should be considered approximates if using equation (4.12). More elaborate simulations of the spectra, e.g. by the program Spinach,[185] is expected to yield better results, but was not performed.

## 4.5 Conclusion

Two new pulse sequences were introduced; the $S^3$ HMBC *homo* and the $S^3$ HMBC *hetero* for the determination of long-range homo- and heteronuclear coupling constants respectively, including size and sign. The pulse sequences were tested on the alkaloid strychnine and resulted in excellent fit of the experimental data to theoretical and reported coupling constants alike.

## 4.6 Experimental

The experimental setup, with relevant information is included in the various tables and figures throughout this chapter and standard pulse sequence setups are found in Appendix A1.

**Acquisition**

For strychnine the setup of the $S^3$ HMBC experiments were: ns = 16, ss = 32, si = 4096, ni = 256 $J_{min}$ = 124 Hz, $J_{max}$ = 169 Hz, $\tau$ = 3.4 ms, $\Delta$ = 65 ms or 62.5 ms.

For vinyl acetate the setup of the $S^3$ HMBC experiments were: ns = 16, ss = 32, si = 4096, ni = 256 $J_{min}$ = 130 Hz, $J_{max}$ = 190 Hz, $\tau$ = 3.1 ms, $\Delta$ = 62.5 ms.

5 mm tubes were used throughout (0.5 mL). For all compounds the $^1J_{CH}$-coupling constants were determined prior to acquisition to input minimum and maximum $^1J_{CH}$-coupling constants into the setup of the $S^3$ HMBC experiments.

**Simulations**

Structures were generated by the modelling suite Maestro version 10.2.010 (2015) by Schrödinger for force field calculations,[126] using the program MacroModel version 10.8.[127,128] The MMFFs force field was used. To generate structures which should cover the conformational space of compounds, a conformational search was performed for each structure by the program MacroModel using energy cutoff of 50 kJ/mol, 100,000 steps and CPRG minimization.[127,129]

Gaussian version 09 revision B.01 was used for DFT calculations including optimizations and NMR calculations.[131] Structures were optimized to a B3LYP/6-31(d) level of theory unless otherwise stated.

# 5 Theory – 3D structural information from NMR, part 2

The following chapters all concern the use of residual dipolar coupling constants to obtain 3D structural information. Theory on the subject is thus presented below.

## 5.1 Residual dipolar coupling

Residual dipolar coupling constants (RDCs) have gradually made the move from macromolecules to small molecules over the last couple of decades. In 2000, Courtieu *et al*. reported enantiodiscriminating properties of poly-γ-benzyl-L-glutamate (PBLG) for the very small molecule 1-chloropropan-2-ol (CP), one of the methods being RDC analysis.[186] In the following years, Thiele and coworkers showed the possibility of solving the pro-chirality of the methylene groups in strychnine, illustrating the possibility of discerning stereoisomers of small molecules.[187,188] Later again, the groups of e.g. Thiele and Reggelin started to investigate the enantiodiscriminating properties of PBLG and other, novel, liquid crystals.[189–192] In parallel, pioneered by the work of Kessler *et al.* in 2004, an investigation into the usage of swollen polymers as alignment media for small molecules began, utilizing an easier scalability of experimental RDCs as well as a weaker degree of alignment.[193,194]

The basis of the residual dipolar couplings is the direct dipolar coupling constant (*D*). Akin to the NOE relaxation, *D*-couplings operate through space via magnetic dipole-dipole interactions from the magnetic fields generated by the spins.[11] In contrast to NOEs and *J*-coupling constants, which are products of interactions between the nuclear spins, dipolar couplings depend on the direct influence of the external magnetic field on internuclear vectors.[11,195,196] *D*-coupling constants depend on the gyromagnetic ratio ($\gamma$) of the involved nuclei, the cubed length of the inter-nuclear vector ($r$) and the angle of said vector to the magnetic field ($\theta$).[V] The other constants are the vacuum permeability constant ($\mu_0$) and the reduced Planck constant ($\hbar$). The relevant formula is given in equations (5.1) and (5.2).[11,195–197]

$$\hat{\mathcal{H}}_{IS}^{DD,full} = -\frac{\mu_0}{8\pi^2}\frac{\hbar\gamma_I\gamma_S}{r_{IS}^3}\left(3\left[\hat{I}_I \cdot e_{IS}\right]\left[\hat{I}_S \cdot e_{IS}\right] - \hat{I}_I \cdot \hat{I}_S\right) \quad \textbf{(5.1)}$$

Where $e_{IS}$ is a vector through the two involved nuclei defined so that $e_{IS}\cdot e_{IS} = 1$. For two spin ½ nuclei and using that the scalar product of the internuclear vector and the magnetic field vector is equal to cos $\theta_{IS}$, equation (5.1) may be rewritten to equation (5.2).[VI]

$$D_{IS} = -\frac{\hbar\gamma_I\gamma_S\mu_0}{16\pi^2}\left\langle\frac{1}{r_{IS}^3}(3\cos^2\theta_{IS} - 1)\right\rangle = -D_{max}\left\langle\frac{1}{r_{IS}^3}(3\cos^2\theta_{IS} - 1)\right\rangle \quad \textbf{(5.2)}$$

In anisotropic liquid conditions the observed angle is an average the dipolar coupling over all orientations. The dependence on the angle to the magnetic field may be used in determining the 3D

[V] Note that $\theta$ changes definition from dihedral angle to angle with magnetic field dependent on whether *J*-coupling constants or RDCs are the current subject.

[VI] Compared to the notation by Levitt, the dipolar coupling is divided by $2\pi$ to gain values in Hz instead to radians/s.

structures of molecules as it will establish angular knowledge of internuclear vectors even far apart in structure, as exemplified in Figure 5.1, which may be utilized by back-calculation and fitting of experimental data to a proposed structure.

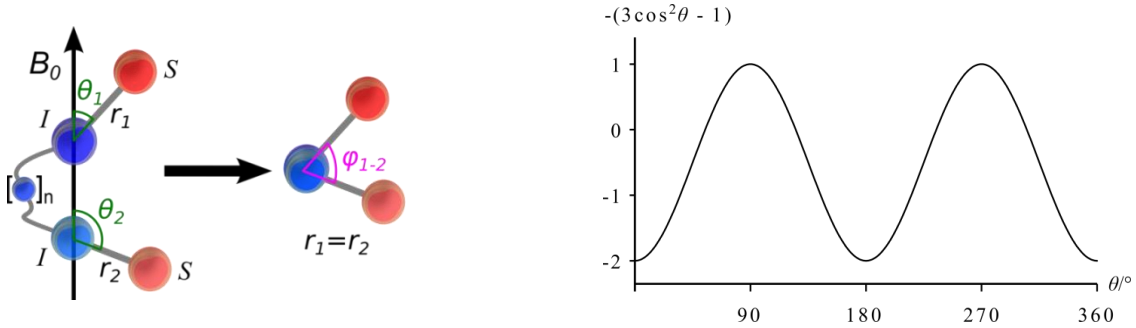



**Figure 5.1. Left: Correlation of the angle between two internuclear vectors ($\varphi$) from the angle of the internuclear vectors to the magnetic field ($\theta$). Right: The relative dependence of RDCs on the angle $\theta$.**

This can make it possible to correlate clusters of stereo-centers in a given structure that are not connected by protons in close proximity, which is impossible using NOEs and *J*-coupling constants due to their local nature.

In isotropic tumbling, i.e. where the molecules are freely moving in solution, the term ($3\cos^2\theta_{IS} - 1$) averages to zero and no RDCs are observed, taking into account that all orientations have equal probability, in equation (5.3).[11]

$$D_{IS} = \int_0^{\pi} \sin\theta_{IS}\,(3\cos^2\theta_{IS} - 1)\,d\theta_{IS} = 0 \qquad \textbf{(5.3)}$$

This implies that *D*-coupling constants may not be observed in isotropic liquid NMR spectroscopy as signal splittings, and a surplus of a given alignment to the magnetic field needs to be introduced making the sample anisotropic. The anisotropic sample is here characterized by a preferred molecular orientation as well as rapid translational motions of the molecules.[11] The result is that intramolecular dipoles give rise to dipolar coupling constants, while intermolecular dipole interactions will still average to zero as a result of translational motions, at least to a good approximation and for short (and thereby larger observed) intermolecular dipole interactions.[11]

Were all molecules equally aligned, the full dipolar coupling would be observed and the constants from equation (5.2) yield *D*-coupling constants with a magnitude of tens of thousands Hz for $^1D_{CH}$ ($D_{max}/r^{-3}$ ~ 11,350 Hz). This would make them hard to measure, and the sign next to impossible to determine, when comparing to $^1J_{CH}$ (typically +100 to +200 Hz).[196,197] Of course other problems such as intermolecular dipolar couplings and broader line-shapes due to a multitude of large long-range dipolar couplings etc. may also be envisioned. Thus the alignment needs to create a smaller surplus of one orientation, with the observed value being the residual dipolar coupling (RDC), preferably in the magnitude of tens of Hz which translates to a ~$10^{-3}$ fraction of alignment. This is in practice achieved by using alignment media in liquid solutions which align only an appropriate amount of molecules leading to average coupling constants of the desired maximum absolute values. Note that *D* will refer to a residual dipolar coupling constant in the rest of this thesis.

In an anisotropic sample the *D*-coupling constants are extracted from the difference between the observed total coupling (*T*) and the *J*-coupling from isotropic tumbling as in equation (5.4).

$$D = T - J \tag{5.4}$$

The sum of dipolar interactions is observed due to identical form of the respective Hamiltonians of the spin parts of *J* and *D*. The *J*-coupling constants are considered unaltered as they are almost independent of the orientation to the magnetic field.[11] The anisotropic part of the *J*-coupling (termed the *J*-anisotropy) is small and usually ignored, even in solids NMR.[11] The line-splitting is thus a summation of the two dipolar coupling contributions, and it should be noted that also $T = J + 2D$ is used in the literature.[195] While this will of course alter the observed RDCs, it will not influence the conclusions from back-calculation of RDCs from 3D structures, since all RDCs are scaled equally.[195] An example of an extraction is illustrated in Figure 5.2. The *J*- and *T*-coupling constants are determined by displacement of peaks to maximum overlay throughout the thesis.

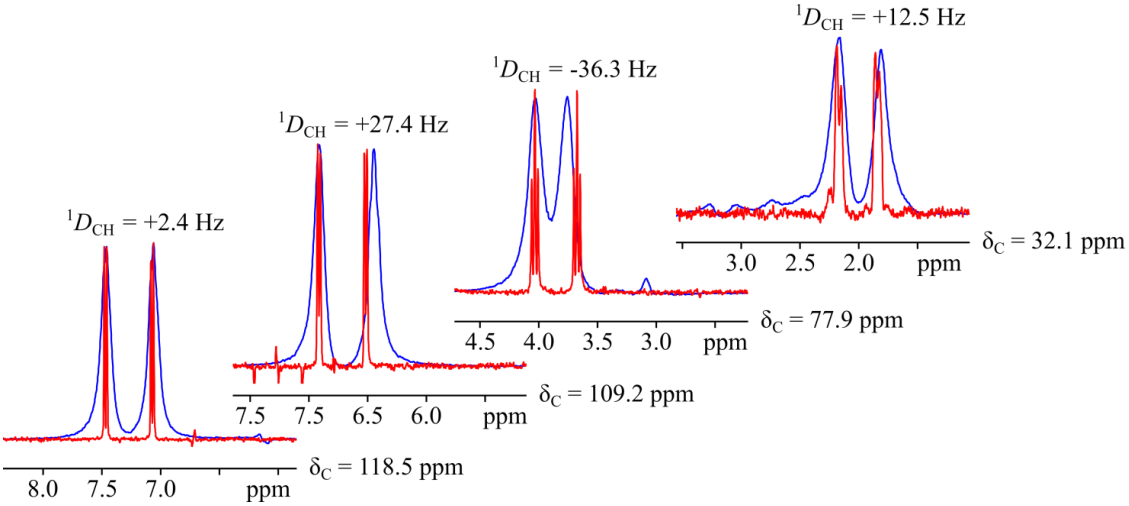



**Figure 5.2. 1D slices of rows ($\delta_C$ indicated) extracted from a CLIP-HSQC of reserpine (see Section 0) under isotropic conditions yielding *J*-coupling constants (red) and anisotropic conditions yielding total coupling constant (blue). Aligned in PDMAA (Polymer 8.2). The values of the RDCs are shown above the doublets.**

To investigate the degree of alignment, the quadrupolar splitting of deuterium nuclei in the deuterated NMR solvents is often used. From the quadrupolar splitting it may be e.g. evaluated whether the alignment is stable across long experiments, and there is a correlation between the quadrupolar splitting and the size of the observed RDCs for many alignment media.[198,199] The linear correlation between the size of RDCs and $^2$H-splitting is compound and alignment media dependent. Not all alignment media give rise to observable quadrupolar splittings, especially the more polar solvents, such as DMSO, seem to exhibit smaller splittings, which may be hidden in the linewidth.[200]

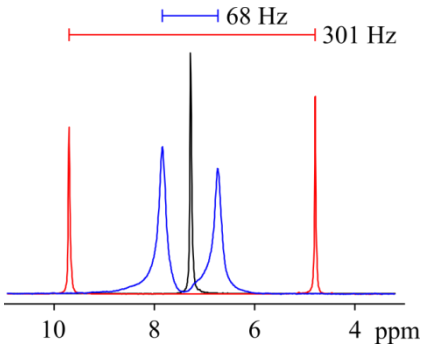



**Figure 5.3. Examples of the $^{2}$H quadrupolar splitting for $CDCl_3$ of an isotropic sample (black), and aligned samples in the liquid crystal PBLG (red) and the stretched polymer PDMAA (blue). The size of the splitting is indicated above at 400 MHz ($^{1}$H frequency).**

The measured RDCs are compared to 3D structures of the investigated molecule by back-calculation, where experimental data is compared to already rendered 3D structure(s). For bio-macromolecules, RDCs are also used in the optimization and simulation of structures, but the method has not caught on within the small molecule community and will not be elaborated on.[201] The preferred back-calculation of RDCs from 3D structures for small molecules is singular value decomposition (SVD),[202–204] which has proven reliable and generally applicable for the structures frequently used in the literature. They generally all share a common feature: They are rigid. The requirement of rigidity in the structure is a major problem for general usage of RDCs, which is addressed in Chapter 0, along with a more detailed introduction to SVD. Other methods have also been proposed such as prediction of alignment based on e.g. shape of the molecule, but these are generally less utilized in the small molecule literature.[203–205]

The used experiment for determining $^{1}J_{CH}$ and $^{1}T_{CH}$ was generally the CLIP experiment, which exhibit the line splitting of the coupling constant in the direct (F2) dimension.[206,207] Other groups have shown that using the indirect (F1) dimension may be advantageous due to e.g. that homonuclear coupling constants will not influence the line splitting in the indirect dimension, resulting in potentially sharper lines.[195,208] The experimental time is increased though due to the lower resolution in F1, even when *J*-scaling is used. The author experimented slight differences in the extracted coupling constants between the F1 and F2 experiments, and it was thought advantageous to keep to one type of experiment throughout.

### 5.1.1 Alignment media

The available alignment media for small molecules in organic solvents are generally divided into two classes: Liquid crystals (LCs) and stretched polymers (SPs).[195,196] The latter class is also termed strain-induced alignment in a gel (SAG) effect alignment media. The amount of alignment media is different compared to those for macromolecules due to a need for applicability in organic solvent. Thus e.g. filamentous phages, rod-shaped cellulose particles or charged acrylamide gels are not generally applicable.[209] The interaction with the alignment media is generally thought to be solely steric, unless a charged alignment media is used.[203,210] A list of alignment media and the solvents in which they may be used is found in Table 5.1.[197]

**Table 5.1. Examples of liquid crystal (LC) and stretched polymers (SP) alignment media and the most utilized solvents used for alignment in the media. Inspired by Thiele[197]**

| LC | Solvent | Chiral | Ref. |
|---|---|---|---|
| PBLG | $CDCl_3$ | Yes | 186,198,211 |
| PELG | $CDCl_3$ | Yes | 188 |
| Guanidines | $CDCl_3$ | Yes | 212 |
| Acetylenes | $CDCl_3$ | Yes | 189,192 |
| Graphene | DMSO | No | 213,214 |

| SP | Solvent | Chiral | Ref. |
|---|---|---|---|
| PS | $CDCl_3$ | No | 193,215 |
| PDMS | $CDCl_3$ | No | 216 |
| PMMA | $CDCl_3$ | No | 194 |
| PVAc | $CDCl_3$ | No | 200 |
| | $CD_3OD$ | | |
| | DMSO | | |
| PH/PAA | $D_2O$ | No | 217,218 |
| | DMSO | | |
| PDMAA/(R/S)-APhES | DMSO | Yes | 210 |
| Gelatin | $D_2O$ | Yes | 219 |

Multiple alignment media are available, and an outline of advantages and limitations of LCs and SPs is given below, as a detailed discussion of all would be beyond the scope of this text. General consideration focused on enantiodiscrimination will also be touched upon as an introduction to Chapter 7. The work concerning enantiodiscrimination of organic molecules has primarily been centered on LCs.[189,191,192,198,212] These include PBLG, polyacetylenes and polyguanidines. For comparison only a single stretched radical-initiated crosslinked polymer has been published,[210] while crosslinked amino acid based stretched polymers such as gelatin have also been shown to possess enantiodiscriminative properties.[219]

### 5.1.2 Liquid Crystals

LCs for the application as small molecular alignment media are typically made of long polymeric single strands that possess a secondary structure. The LCs are normally helical but sheet structures have been reported in the form of graphene layers.[198,213] The most prominent member of the class is arguably PBLG, and this is also the only LC the author has worked with.

The anisotropy is introduced by dissolving of the LC in the given solvent.[190,198,220] By repeatedly transferring the solution from one end of the NMR tube to the other, e.g. by use of rotor, the solution is homogenized. The homogeneity along the length of the NMR tube is verified by measuring the quadrupolar splitting of the deuterium signal. The anisotropy is induced by interactions of the molecules and the vertically aligned helices as illustrated in Figure 5.4.

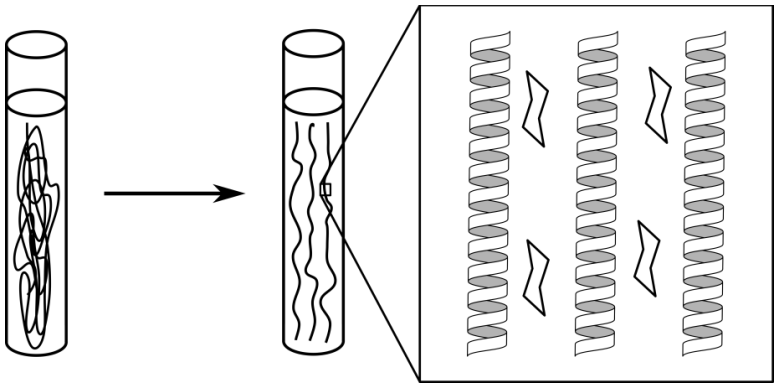


**Figure 5.4. Illustration of the alignment of molecules in liquid crystals (LC), based on PBLG. The induced anisotropy is indicated.**

The main disadvantage of the LCs is that they are hardly scalable, though different molecular weight PBLG was shown to result in differences in the alignment strength.[198] Also the alignment is strong, leading to larger RDCs, which may lead to problems and limit the scope of usage.[198] An advantage is that many LCs are readily available from vendors, the sample preparation of the alignment media is easy and the line shapes of aligned compounds are, based on experimental experience with PBLG of the author, sharper than when using stretched polymers. Also, most of the LCs are inherently chiral and enantiomeric pure, leading to better discrimination between enantiomers.[189,192,198,212]

While PBLG is probably the most utilized LC for RDC extraction, the enantiodiscriminative prowess of the polymer is not great compared to other reported LCs. It was shown though, that the alignment of (+)- and (-)-IPC, and thus the observed RDCs, was affected by the amount of PBLG versus PBDG (the enantiomeric polymer), and that differentiation was indeed possible.[190,198] Also cross-linked PBLG (which could in theory be considered a stretched polymer) was shown to differentiate IPC.[191] It was actually reported that the differentiation was increased by crosslinking the LC.

Other liquid crystals have been published with greater enantiodiscriminative properties. The first was a polyguanidine which showed much differences in the orientation of the enantiomers of IPC.[212] This was followed by two polyacetylenes which also showed great differentiation.[189,192] Interestingly, one of the polyacetylenes was reported to exhibit enantiodiscriminative properties even after the temperature was varied to a point where circular dichroism showed no left- or right-handedness of the helical backbone, suggesting that the enantiodiscriminative properties may in part be due to side-chain chirality.[192] This was surprising since the enantiodiscriminative properties of the LCs were thought to originate mostly from their helical structure, while this finding suggest an interaction to the chiral center presented at the side-chain.[189,190,192,212]

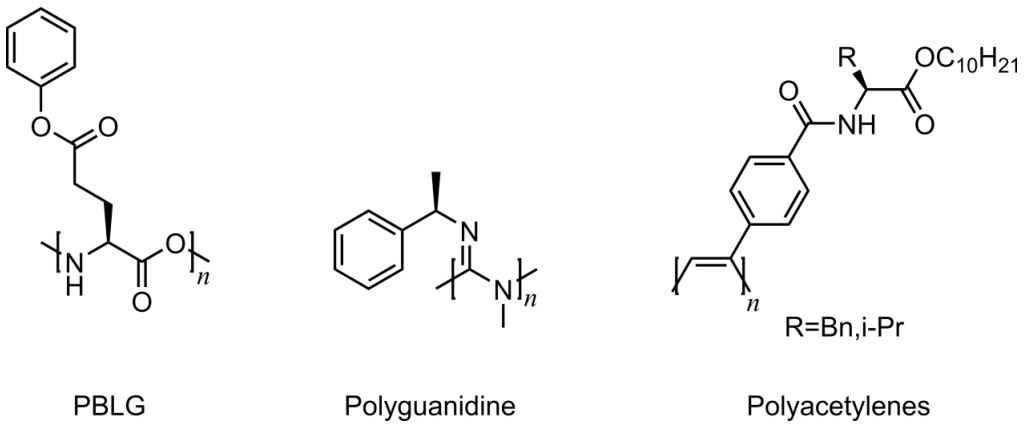



**Figure 5.5. Examples of chiral LCs from the literature. All give rise to helixes in solution.[189,190,192,198,212]**

### 5.1.3 Stretched Polymers

Stretched polymers cover crosslinked polymers that swell in the used solvents. The term stretched polymers is fitting especially for polymers which are mechanically scaled using an amply named stretching apparatus.[199,221–223] The stretching apparatus consists of a rubber tube and clams which

uniformly stretches the gel inside the tube, and may be used to tune the alignment. Rubber tubes for different solvents have been introduced to allow the usage of e.g. chloroform.[199]

Another method of swelling the polymer is by synthesizing polymers of a size which limits swelling to be either radially or vertically in the NMR tube. The polymer stick is then swollen in a suitable solvent and the result is an anisotropic environment as illustrated in Figure 5.6.[194,200] The anisotropic environment may be simplified to the steric interactions with the irregular distribution of interaction surfaces with either horizontal or vertical surface in excess, formed once the polymer is swollen.[194] In practice, the polymers were synthesized polymers in either 5 mm (vertical swelling) or 3 mm (radial swelling) NMR tubes, and subsequently swollen in 5 mm NMR tubes.[194]

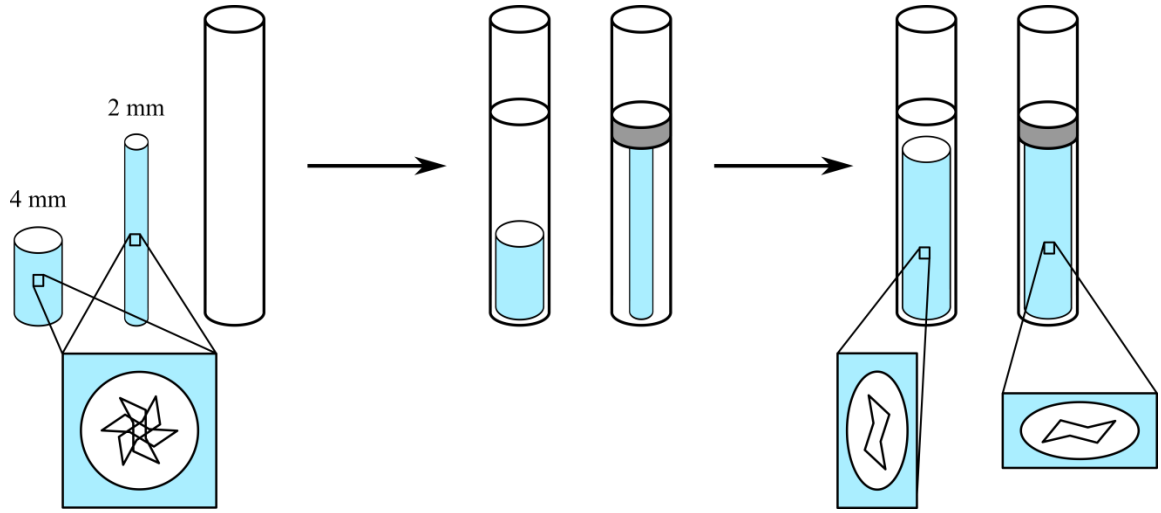
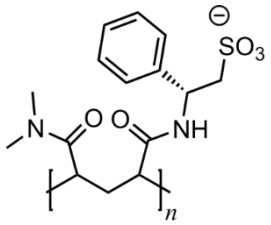



**Figure 5.6. Schematic representation of the alignment of molecules in stretched polymers, either swollen vertically or radially in an appropriate solvent. The anisotropy of the polymers upon swelling is indicated.**

The alignment strength may often not be altered once swollen, but may be tuned by the amount of cross-linker used during synthesis and the size of chosen polymer stick.[194,200] It should be noted that the synthesis of these polymers seems much easier than that of e.g. PBLG (though the latter was not attempted by the author),[198] and it is hardly problematic that scaling depends on cross-linker content. The vertical swelling was first introduced, but the time needed to reach full swelling is generally long, up to several weeks.[194] Thus the concept of radial swelling of thin polymers which were vertically restricted was introduced.[194] Some of the polymers may be shifted from constricted to un-constricted multiple times while other do not possess this property and should not be relaxed after swelling.[194,210]

The amount of stretched polymers which will differentiate enantiomers is, as mentioned previously, lower than the number for LCs. Polymers are based on crosslinked amino acids, such as gelatin and the underlying biopolymer of gelatin, collagen, were shown to differentiate diastereomers in aqueous solution.[219,223,224] A more classic ( i.e. radical initiated) polymer based on a chiral monomer alike to the achiral AMPS monomer, (*R*/*S*)-2-acrylamido-2-methylpropane-1-sulfonic acid (APhES), and the achiral monomer (*N*,*N*)-dimethylacrylamide (1:1) was later introduced, which was shown to differentiate compounds that possess a charged amine group under acidic conditions in DMSO.[210,217]




(R)-APhES/DMAA

**Figure 5.7. Example of a chiral stretched polymer from the literature. Was showed to work only by interactions to a charged amine moiety of the analyte.[210]**

### 5.1.4 Evaluation of RDCs

To evaluate the experimental RDCs in comparison to the given 3D structure(s), the *Q*-factor is utilized as given in equation (5.5).[203,204]

$$Q = \sqrt{\frac{\sum\left(D_i^{exp} - D_i^{calc}\right)^2}{\sum D_i^{exp\,2}}} \qquad \textbf{(5.5)}$$

The *Q*-factor approaches zero when the difference between experimental and back-calculated values approaches zero and the size of the extracted RDCs is included in the calculations, which leads to easy comparison of fits between dataset of different degrees of alignment.

### 5.1.5 Parameters used to investigate enantiodiscrimination

Some different parameters have been proposed in order to compare the alignment of different enantiomers. One is the 5D or generalized cosine β (GCB) angle, which directly compare the alignment tensors from e.g. SVD fittings, given in equation (5.6) as the normalized scalar product between the alignment tensors (see Section 8.1 for the theory of SVD and alignment tensors).[192,225]

$$GCB = \frac{\langle A^1|A^2\rangle}{\langle A^1|A^1\rangle^{½}\langle A^2|A^2\rangle^{½}}$$
$$\langle A^1|A^2\rangle = \sum_{i,j} A_{ij}^1 A_{ij}^2 \qquad \textbf{(5.6)}$$

An F-test has also been utilized to check whether the actual RDC values are statistically different, using the null hypothesis that there is no significant difference (standard for the F-test).[210] The F-test is used to compare variance of the data and the difference in variance between datasets of different and identical enantiomers are compared. It should be noted that the F-test is generally quite sensitive toward non-normality and it is not used in this thesis.

# 6 Determination of long-range residual dipolar coupling constants

A major hindrance in the scope of utilizing RDCs for structural studies of small molecules is an inherent lack of internuclear vectors to measure and compare to 3D structures. Whereas $^1D_{XH}$ are plentiful for macromolecules, only utilizing the one bond couplings is a limitation for small molecule RDCs. Thus it would be beneficial to increase the number of vectors by including long-range RDCs. For this purpose, the $S^3$ HMBC experiments may prove very useful. While SVD is reviewed in Chapter 0, it is worth mentioning that it is the most widely used back-calculation method used for small organic molecules, and that a minimum of five linearly independent internuclear vectors are needed to obtain a fit between RDC data and a rigid 3D structure. If multiple moving parts of a structure need to be investigated independently, five independent internuclear vectors are needed *per moving part*, which may be a concern for small molecules.[202,226]

## 6.1 Utilized alignment media

It was not straight forward to implement the usage of the $S^3$ HMBC experiments in RDCs due to inherent problems with the used alignment media. This may be attributed to the change in $T_2$ relaxation for the small molecules in aligned samples. A decrease in $T_2$ time due to interactions to the solid-like alignment media leads to a rapid loss of magnetization and the longer, compared to e.g. HSQC-type experiments, $S^3$ HMBC experiments resulted in a lack of signal to detect. Thus the search for an alignment medium which proved advantageous was initiated. Essentially two factors were deemed the most important:

1. Alignment should be weak enough to allow sufficient tumbling so that $T_2$ relaxation allows signal detection
2. Alignment strong should enough that long-range RDCs are larger than the experimental error

As one might suspect, these requirements are contradicting and some alignment media may be readily discarded.

The initial choice of alignment media illustrates this dilemma; the liquid crystal PBLG and stretched PMMA and PVAc polymers.[194,198,200] Spectra of IPC and strychnine from PBLG failed to obtain an *S*/*N* ratio where RDCs could be determined in the author's hands. This is contributed to the strong alignment of the medium, perhaps best illustrated by the large observed $^1D_{CH}$-coupling constants. On the contrary, PMMA and pVAc would give rise to observable RDCs, but weak alignment strength and thus low RDCs, gave rise to a large possible influence from experimental errors and low discriminative prowess. This was somewhat relieved by altering the amount of cross-linker in order to increase the size of the RDCs.

In the end, an alignment medium that best satisfied the needs was identified; a new stretched polymer made from crosslinked DMAA (PDMAA). It should be noted that a similar polymer has been published, e.g. the PH/PAA polymer where 2-(acrylamido)-2-methylpropanesulfonic acid is used in addition to DMAA (in a molar fraction of 1:1) and crosslinker.[210,217] This is supposedly done to facilitate swelling in DMSO.

It was advantageous to swell PDMAA using the confinement method, where vertical swelling is restricted and the polymer is allowed to swell radially. PDMAA swell easily in both chloroform and DMSO with observed coupling constants in the order of 20-30 Hz, and is thus very well suited for the purpose of aligning small organic molecules. The observed RDCs were easily scalable by differentiation of the amount of cross-linker in line with other stretched polymers.[194] This is advantageous, native to stretched polymers over LC, as the optimal alignment of different compounds was found by a singular polymer constitution, and that the alignment strength differed between compounds in the same polymer, even when comparable structures were aligned.

## 6.2 Results - $S^3$ HMBC *homo*

The $S^3$ HMBC *hetero* experiment was a relatively late addition to the work, and was finalized while writing the thesis. Thus, long-range heteronuclear results from aligned samples are not included and this section will focus solely on $S^3$ HMBC *homo*.

RDCs were extracted from spectra of aligned and isotropic samples of strychnine and IPC. The basis of the alignment tensors for the structures was $^1D_{CH}$-coupling constants from CLIP-HSQC spectra, since the long-range $^nD_{HH}$-coupling constants are inherently smaller than $^1D_{CH}$-coupling constant due to the $r^{-3}$ relationship. While much focus has been on measuring long-range heteronuclear coupling constants in the case of *J*-coupling constants,[145,146] long-range homonuclear RDCs are more interesting than their hetereonuclear counterparts due to the ~4:1 ratio of the gyromagnetic ratios of $^1$H and $^{13}$C, as illustrated in Table 6.1. As a result quite long distances between protons have identical potential values as the shortest of long-range heteronuclear RDCs ($^2D_{CH}$). While examples of homonuclear RDCs are available in the literature, they generally seem sparsely used.[227–229]

**Table 6.1. Comparison of theoretical relative *D*-coupling constants to that of $^1D_{CH}$ for different distances (*r*) calculated using equation (5.2). Identical angle to the magnetic field is assumed. Examples of nuclei pairs with an internuclear vector of the used distances are given.**

| *r* [Å] | $D_{CH}$ | $D_{HH}$ | Example |
|---|---|---|---|
| 1.1 | 1.00 | - | $r(^1D_{CH})$ |
| 1.8 | 0.24 | 0.94 | $r(^2D_{HH})$ |
| 2.2 | 0.13 | 0.50 | $r(^2D_{CH})$ |
| 2.6 | 0.06 | 0.24 | - |
| 3.0 | 0.04 | 0.15 | - |

The number of RDCs obtained from $S^3$ HMBC *homo* was generally lower than the number of *J*-coupling constants extracted from isotropic samples, possibly due to an increase in $T_2$ relaxation in the anisotropic samples. Another explanation could be altered long-range heteronuclear coupling constants leading to a too low or high coupling constants compared to the length of the delay used in the pulse sequence. The latter might possibly be alleviated by using an array of delays, but this was not tested.

### 6.2.1 IPC

The structure of IPC is much used, probably due to favorable rather weak alignment in most media, the rigid structural scaffold and multiple non-parallel vectors.[189,190]

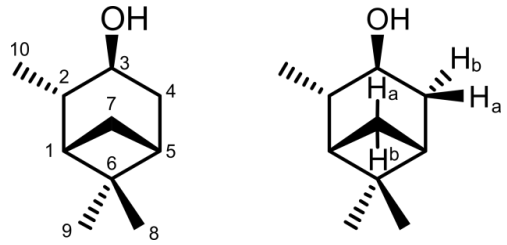



**Figure 6.1. Structure of (+)-IPC. Stereochemistry of chiral centers and diastereotopic protons indicated.**

IPC was aligned in PMMA, using the confinement method to introduce anisotropy. The measured one-bond and long-range RDCs are given in Table 6.2. Note that most of the "long-range" coupling constants were in fact $^{3}D_{HH}$, which could be obtained from other experiments, though the $S^3$ HMBC *homo* provide an accurate and easy method of performing the measurements. Also some $^{4}D_{HH}$ were extracted. The correlation of the data is generally good, though the *Q*-factor increased upon including the extra RDC data.

**Table 6.2. Experimental and back-calculated RDCs of IPC using $^{1}D_{CH}$ from CLIP-HSQC only and including $^{n}D_{HH}$ from $S^3$ HMBC *homo*. When no experimental data (-) is given the theoretical values are calculated from the existing data. Aligned in PMMA (Polymer 6.1)**

| Nuc 1 | Nuc 2 | Exp [Hz] | Comp [Hz] | Exp [Hz] | Comp [Hz] |
|---|---|---|---|---|---|
| C1 | H1 | 16.3 | 16.7 | 16.3 | 18.2 |
| C2 | H2 | -11.6 | -9.6 | -11.6 | -8.8 |
| C3 | H3 | 14.6 | 14.2 | 14.6 | 12.5 |
| C4 | H4a | -4.8 | -4.8 | -4.8 | -6.1 |
| C4 | H4b | 15.5 | 14.2 | 15.5 | 16.1 |
| C5 | H6 | -7.4 | -7.8 | -7.4 | -5.8 |
| C7 | H7a | -21.7 | -23.2 | -21.7 | -23.8 |
| C7 | H7b | 4.9 | 5.2 | 4.9 | 2.7 |
| H1 | H6 | - | 0.7 | 1.0 | 0.9 |
| H1 | H7a | - | 0.2 | 0.5 | 0.0 |
| H2 | H3 | - | 3.7 | 3.7 | 4.0 |
| H2 | H4b | - | 0.5 | 1.5 | 0.8 |
| H3 | H4a | - | 1.4 | 1.6 | 1.1 |
| H3 | H4b | - | -1.7 | -2.1 | -1.0 |
| H3 | H6 | - | -0.4 | -0.3 | -0.5 |
| H4a | H6 | - | -7.5 | -7.0 | -7.9 |
| | | *Q*-factor | 0.076 | | 0.147 |

**Table 6.3. *Q*-factors of back-calculations of IPC of correct assignment and with switched chirality of diastereotopic protons.**

| Switched | *Q*-factor | | | |
|---|---|---|---|---|
| | $^{1}D_{CH}$ | Ratio | $^{1}D_{CH}$+$^{n}D_{HH}$ | Ratio |
| Right | 0.076 | 1.0 | 0.147 | 1.0 |
| Dia-2 | 0.661 | 8.7 | 0.806 | 5.5 |
| Dia-3 | 0.305 | 4.0 | 0.432 | 2.9 |
| Dia-4 | 0.201 | 2.6 | 0.662 | 4.5 |
| Dia-7 | 0.659 | 8.7 | 0.668 | 4.5 |
| Dia-2+3 | 0.655 | 8.6 | 0.759 | 5.2 |
| Dia-4+7 | 0.440 | 5.8 | 0.678 | 4.6 |
| Dia-2+4+7 | 0.284 | 3.7 | 0.339 | 2.3 |
| Dia-2+3+4+7 | 0.283 | 3.7 | 0.375 | 2.6 |

The additional information from the homonuclear coupling constants was used in an attempt to determine the diastereomer of the known structure. The result is found in Table 6.3 and Figure 6.2, and the additional information led to little or no differentiation compared to only using the one bond coupling constants, as the lower values of the long-range RDCs correlate well to all 3D structures.

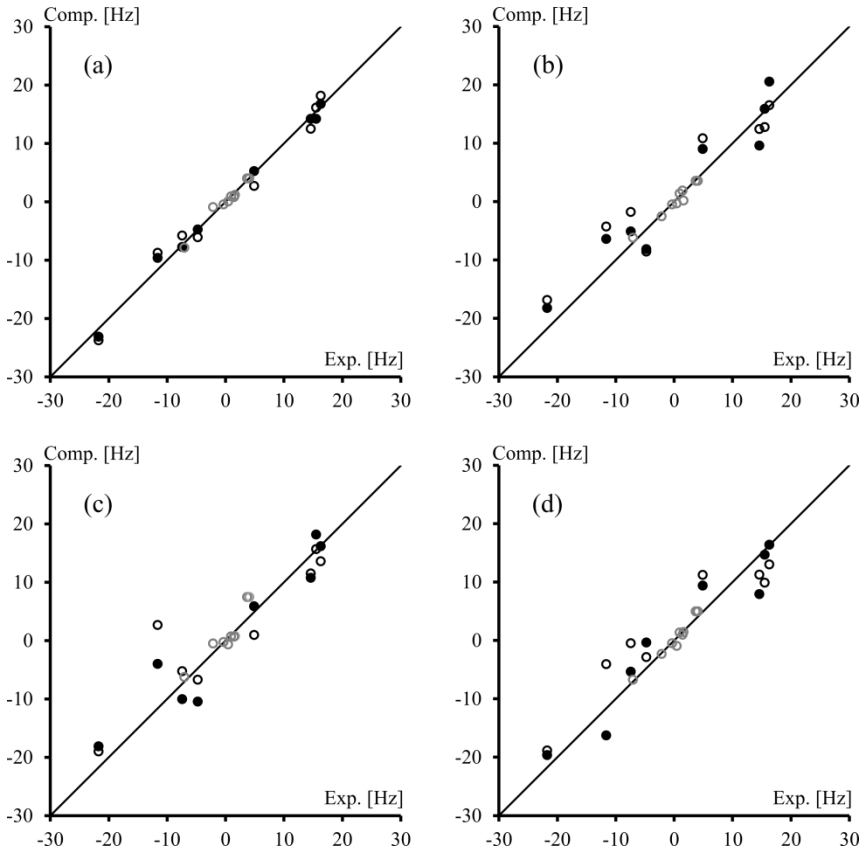



**Figure 6.2. Comparison of experimental and back-calculated RDCs of IPC. Black dots: only $^{1}D_{CH}$ data. Open circles: $^{1}D_{CH}$ and $^{n}D_{HH}$ data used, with $^{1}D_{CH}$ in black and $^{n}D_{HH}$ in grey. Aligned in PMMA (Polymer 6.1)**

**a) Correct stereochemistry.**

**b) 2-stereoisomer.**

**c) 3-stereoisomer.**

**d) 2- and 3-stereoisomer.**

## 6.2.2 Strychnine

Strychnine was also utilized to showcase the scope of the new $S^3$ experiments. The alignment medium was PDMAA which was identified to result in the highest number of coupling constants from $S^3$ HMBC *homo* versus alignment strength ratio. While PMMA gave rise to similar magnitude $^{1}D_{CH}$-coupling constants as DMAA, only a few $^{n}D_{HH}$-coupling constants could be extracted from the $S^3$ HMBC *homo* spectra in PMMA. Examples of extracted RDCs are given in Figure 6.3.

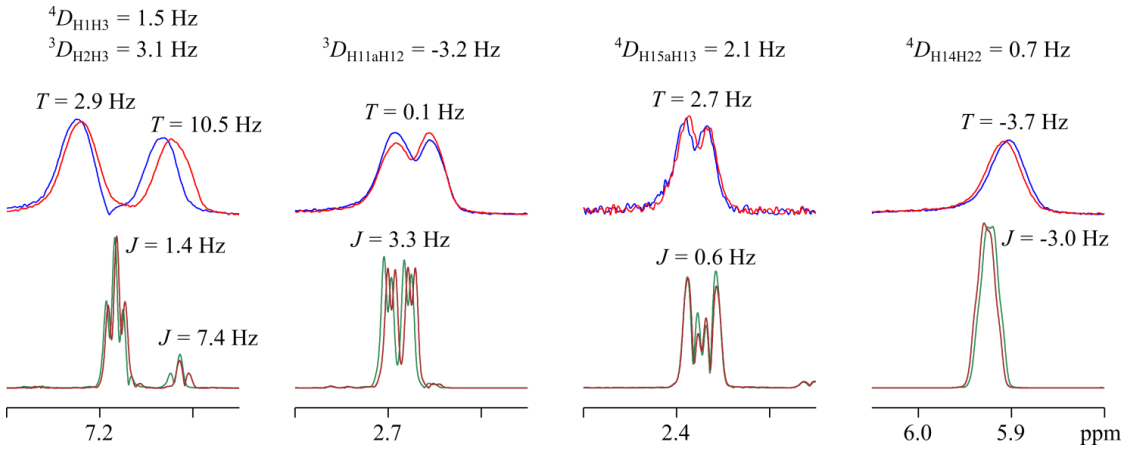



**Figure 6.3. Comparison of $S^3$ HMBC *homo* spectra of strychnine in $CDCl_3$ under isotropic (green and brown) and anisotropic (blue and red) conditions. Aligned by PDMAA (Polymer 6.2). The extracted coupling constants and associated RDCs are given.**

The general findings were similar to those of IPC, with the addition of the extra coupling constants leading to a slightly worse *Q*-factor. Many long-range coupling constants were extracted, more than doubling the available number of internuclear vectors included in the back-calculations of the data (26 $^{n}D_{HH}$ in addition to 18 $^{1}D_{CH}$) which is illustrated in Figure 6.4, values given in Appendix A4. While the extra internuclear vectors were not needed in the fitting of strychnine, it showed the potential of using the $S^3$ HMBC experiments in extracting RDCs.

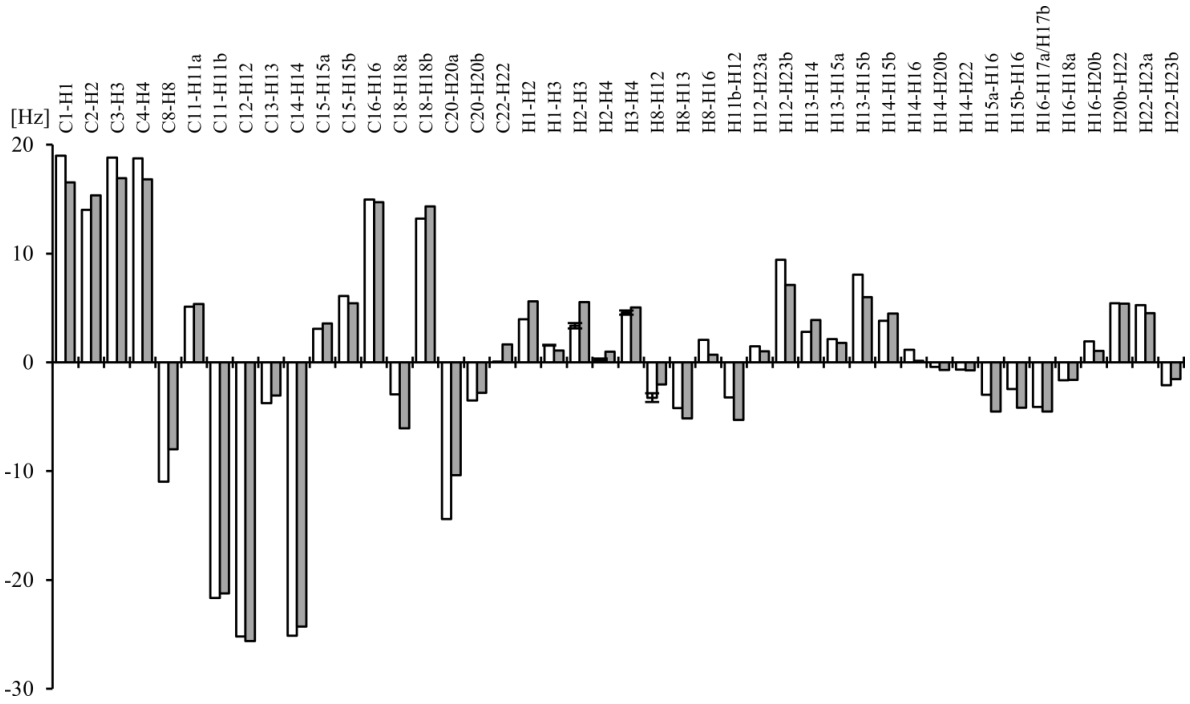



**Figure 6.4. A graphical representation of the experimental (white) and back-calculated (grey) RDCs of strychnine. When RDCs were extracted from two directions the average is given and the difference indicated with error bars. Anisotropy induced by PDMAA (Polymer 6.2).**

When comparing experimental data to back-calculated values from wrongfully assigned diastereotopic protons as seen in Figure 6.5, the best correlation between experimental and back-calculated RDCs was the correct assignment, and only the diastereomers of 15a/15b was virtually indistinguishable due to very similar RDCs. The stereochemistry of the diastereotopic protons 23a/23b could not be solved without the additional data from $^{n}D_{HH}$-coupling constants due to overlap of the resonances in the CLIP-HSQC. This is alleviated by the usage of add and subtract spectra in the $S^3$ HMBC *homo* as the coupling constants are not in-phase and thus the resonances are more easily differentiated. A similar distinction might be possible using the CLIP/CLAP-HSQC approach.[206] The differentiation of the diastereotopic protons was facilitated by a favorable (and lucky) orientation of the aligned molecules, where the internuclear vectors that include H23a/b had rather large and different RDCs compared to the remaining long-range RDCs.

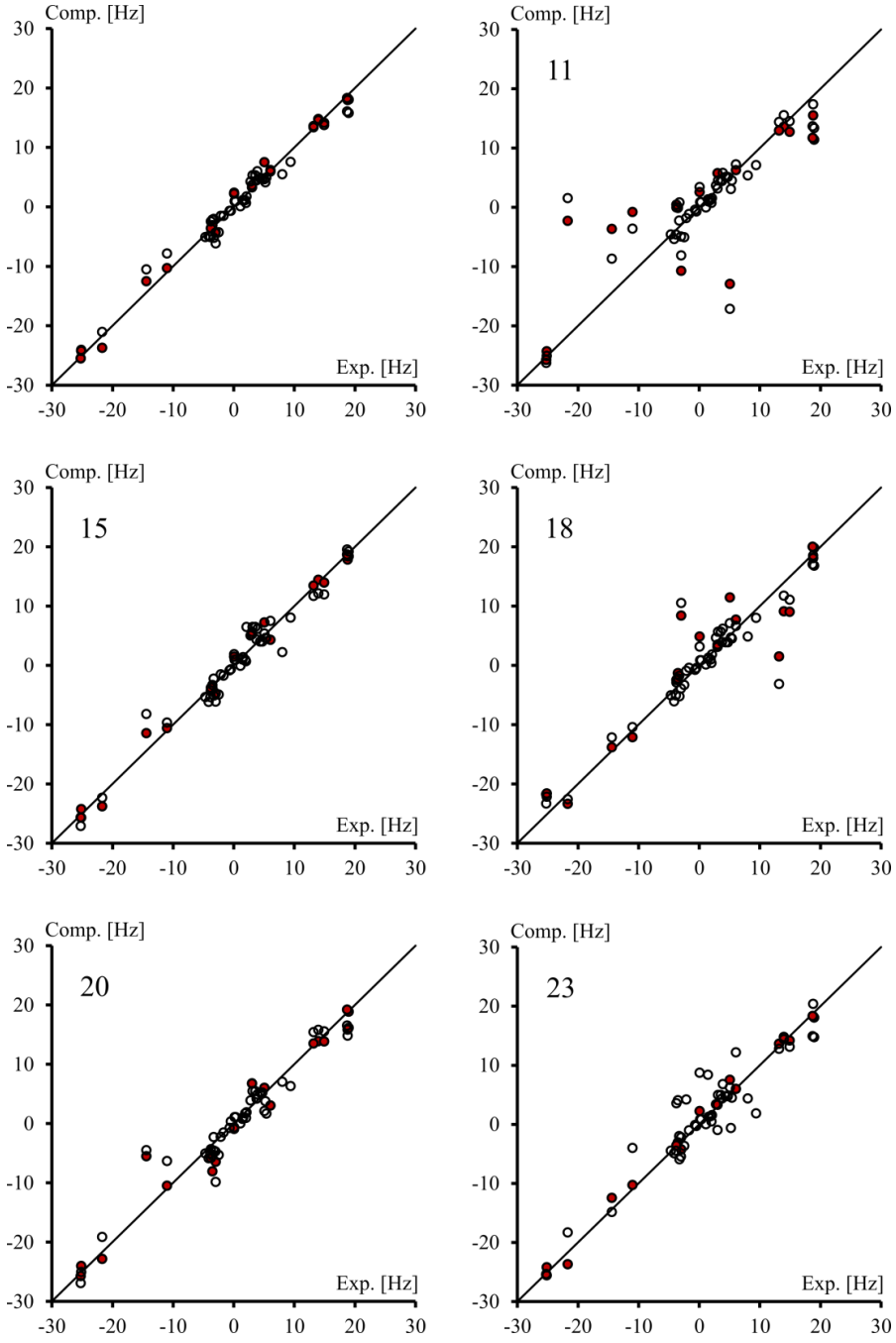



**Figure 6.5. Comparison of experimental and back-calculated RDCs of strychnine. Red dots: only $^{1}D_{CH}$ data. Open circles: $^{1}D_{CH}$ and $^{n}D_{HH}$ from $S^3$ HMBC *homo*. Upper left is right stereochemistry, and the diastereotopic protons are switched in the rest as indicated in the plots. Anisotropy induced by PDMAA (Polymer 6.2).**

**Tables with the data are found in Appendix A4.**

## 6.3 Conclusion and perspectives

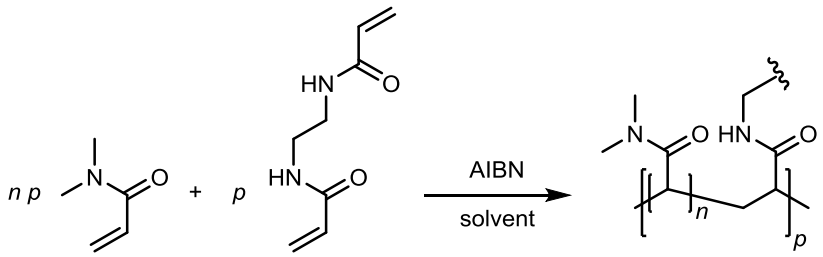
The $S^3$ HMBC *homo* may be used to extract long-range homonuclear RDCs accurately and, if aligned by a suitable alignment media, in a large number. It should of course be noted here, that the chosen compounds both have a high proton to carbon ratio. The increase in available internuclear vectors may greatly impact for small molecular NMR spectroscopy, where the number of one-bond RDCs is often limited. The length of the pulse sequence may be a limitation, presumably due to a faster $T_2$ relaxation induced by the alignment media, but alignment media were identified which enable extraction of the total coupling constants.

## 6.4 Experimental

**Synthesis of achiral polymers**

The syntheses presented are generally inspired from the literature.[192,194,210,215] All commercially bought monomers and crosslinkers (except *N*,*N*-methylenbisacrylamide which is a solid) were purified prior to the synthesis to remove the polymerization inhibitor by passing the neat liquid through a pipette filled with basic alumina.

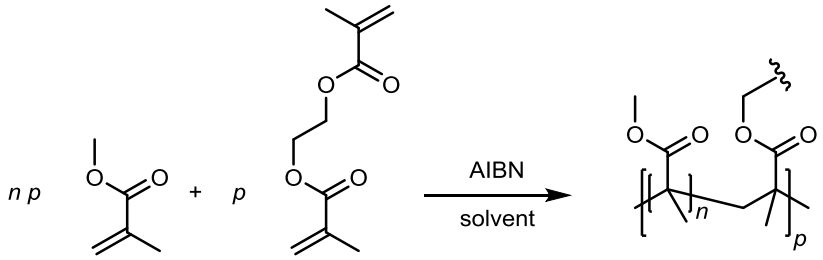
**PDMAA**




To 1 mL of *N*,*N*-dimethyl acrylamide (9.7 mmol) was added 1.6 mg (0.009 mmol) *N*,*N*-methylenbisacrylamide and 1.5 mg azobisisobutyronitrile (AIBN) (0.009 mmol). $N_2$ was bubbled through the solution for 5 minutes to remove $O_2$ and the solution was transferred to 3 mm (or 5 mm) NMR tubes. Air was removed under vacuum and $N_2$ backfilled into the tubes three times and the tubes were left at 60 °C overnight (approx. 14-16 h) to polymerize.

A solvent ($CDCl_3$, DCM, acetone-$d_6$) may be added before transfer, and was usually used in 1:2 ratio of solvent:*N*,*N*-dimethyl acrylamide (*v:v*).

**PMMA**




The protocol of Gil *et al.* was generally followed.[194]

10 mL MMA (61.0 mmol), was added 2 mL acetone-$d_6$ and 2.5 mg AIBN (0.015 mmol). 10 mL were taken out and mixed with 40 µL ethylene glycol dimethylacrylate (EGDMA) (0.42 mmol). The solution was transferred to 3 or 5 mm NMR tubes evacuated and back-filled with $N_2$ three times. The cycle was repeated 3 times, and the tubes were left at 50 °C overnight (approx. 14-16 h).

**Swelling of polymers**

**5 mm sticks**

The tubes were left for a couple of hours without lid after being cooled to room temperature. The polymers could usually be removed easily from the tubes without need of breakage. If stuck the tubes were carefully broken though. The polymers were then cut to appropriate length (2 cm) and

inserted into a 5 mm NMR tube. The solvent and analyte was added and the polymer was left to swell. Depending on the polymer the swelling time varied from approximately one week to several weeks. The degree of alignment was carefully monitored from the quadrupolar splitting of the deuterated solvent.

**3 mm sticks**

The tubes were carefully broken and left overnight to dry. This was especially beneficial for the more sticky chiral polymers. The glass was carefully removed from the polymers and the resulting polymer sticks were cut into the appropriate length (4 cm). The cut polymer stick was inserted into a 5 mm NMR tube and the analyte was dissolved in 0.3 mL solvent and added. A plug was quickly inserted to prevent vertical swelling. Samples should be left for at least 5 h ($CDCl_3$), 1 d (DMSO-$d_6$) or 2-3 d (MeOD) for swelling. To make sure that complete swelling was achieved, a minimum 1d ($CDCl_3$), 3 d (DMSO-$d_6$) or 5 d (MeOD) was always used.

In the table below the constituents in the synthesis of the relevant polymers are given.

| Name | Analyte (c [mM]) | Solvent | Monomer | Crosslinker (mol%) | Monomer:solvent in synthesis (v:v) | Field [MHz] |
|---|---|---|---|---|---|---|
| Polymer 6.1 | IPC (553.6) | CDCl3 | PMMA | EGDMA (0.38) | 4.5:1 acetone-$d_6$ | 800 |
| Polymer 6.2 | Strychnine (216.5) | CDCl3 | DMAA | MBAA (0.34) | - | 800 |

**NMR experiments**

The CLIP-HSQC experiments was used for the determination of one-bond CH coupling constants, using standard setup as described in Appendix A1. Isotropic spectra of IPC (145 mM) and strychnine (144 mM) were acquired in 5 mm tube (0.5 mL)

The $S^3$ HMBC *homo* was used for long-range homonuclear coupling constants as described in Chapter 4, using standard setup as described in Appendix A1.

Total coupling constants were extracted from aligned samples, while *J*-coupling constants were extracted from isotropic samples, in order to determine the RDCs, as described in Section 5.1.

**Simulations**

Structures were generated by the modelling suite Maestro version 10.2.010 (2015) by Schrödinger for force field calculations,[126] using the program MacroModel version 10.8.[127,128] The MMFFs force field was used. To generate structures which should cover the conformational space of compounds, a conformational search was performed for each structure by the program MacroModel using energy cutoff of 50 kJ/mol, 100,000 steps and CPRG minimization.[127,129]

Gaussian version 09 revision B.01 was used for DFT calculations including optimizations and NMR calculations.[131] Structures were optimized to a B3LYP/6-31(d) level of theory unless otherwise stated.

# 7 Chiral alignment media for enantiodiscrimination

Nature is chiral – and most biological molecules are chiral structures.[230,231] Usually one enantiomer of a given macromolecule or small molecule is universally present in an organism, leading to highly specific chiral environments. In line with the inherent chirality of nature, the chirality of target molecules in organic synthesis, as well as molecules isolated from natural sources, is of utmost importance. This is a consequence of the compounds often being interesting due to interactions with a natural target.[230,231] The practical importance may be exemplified by a drug hoping to reach the US market. To be qualified for approval, the FDA demands that the compound is enantiomeric pure, or that both enantiomers are evaluated.[232] This is due to often very different pharmacological properties of enantiomers caused by the chiral environment of the human body.[230]

While chiral centers may usually be solved from a single known stereocenter, and diastereomers thus may be separated, enantiomeric structures are much harder to resolve. Similar physical properties of the enantiomers result in identical observables in the achiral environments where small molecules are usually investigated, be that utilizing MS, IR, UV or NMR methods. One of the major selling points for RDCs in small molecular NMR, besides providing an orthogonal method to e.g. NOE connectivities and *J*-coupling constants in obtaining 3D structural information, is a promise of differentiation of enantiomers. This is arguably the most challenging stereochemical problem to solve in organic as well as natural product chemistry, and thus a premium goal for NMR spectroscopy.

## 7.1 Methods of enantiodiscrimination

Due to the important functional differences of enantiomers, the differentiation of these has been a major research topic for years. RDCs are a relatively new approach, and other methods will thus briefly be mentioned. While NMR spectroscopy is utilized in some of the methods, many utilize different analytical tools.

The probably most utilized method of discerning enantiomers is by optical rotation, where it is utilized that the enantiomers will rotate plane-polarized light in opposite directions.[233] The analysis is carried out by shining single-wavelength, plane-polarized light through a solution of the analyte and the rotation is measured; a negative value means a rotation to the left, and a positive to the right. The specific rotation $[\alpha]$ is given by equation (2.1), at 20 °C and using a sodium lamp of wavelength 589 nm.

$$[\alpha]_D^{20} = \frac{\alpha}{c \cdot l} \qquad \textbf{(7.1)}$$

Where $\alpha$ is the observed rotation, $c$ is the concentration of the analyte and $l$ is the path length of the light. The rotation is dependent on chirality as well as other structural features and a reference of known chirality is needed to determine the absolute stereochemistry.[233]

Another widespread method to distinguish enantiomer is by chemical derivatization, where a specific functional group is derivatized using a chiral building block of known absolute stereochemistry. The derivatization changes the enantiomers to diastereomers, which may be separated by e.g. standard HPLC systems, an may lead to different resonances in NMR spectra. A

multitude of reagents are available dependent on the nature of the compound which is investigated.[234–238] Some of the more well-known include Mosher's reagent for, primarily, derivatization of secondary alcohols and Marfey's reagent for amino acids.[82,83] Using Marfey's reagent as an example, the diastereomers are compared to diastereomers of known chirality, synthesized from pure amino acids, as illustrated in Figure 7.1. The subsequent separation on a HPLC column obtained by Marfey's reagent is increased by steric differences in the structures resulting in stronger intramolecular hydrogen bonds of the L-isomer.[82]

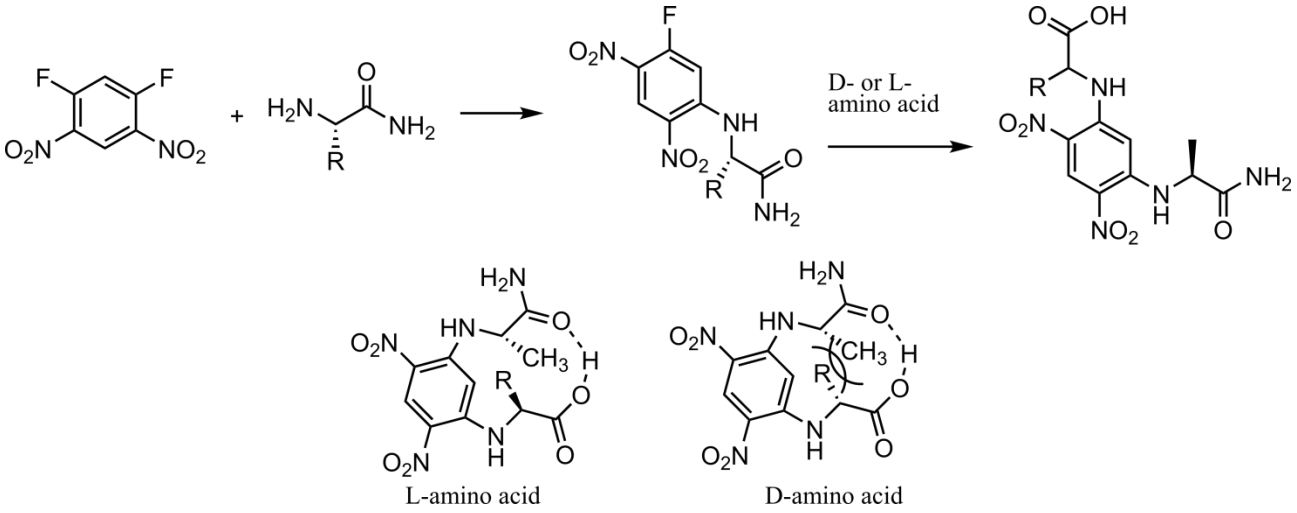



**Figure 7.1. Top: Marfey's reagent, made from 1,5-Difluoro-2,4-dinitrobenzene and L-alanine-NH2 which is reacted with an amino acid. Bottom: The difference in the spatial properties of the complex when reacted with L- or D-amino acids.**

Of methods to differentiate enantiomers by NMR other than RDCs, paramagnetic lanthanide chemical shift reagents and chiral solvating agents should be mentioned.[237,238] The lanthanide reagents are complexes of enantiomerically pure organic structures, e.g. based on camphor derivatives, coordinated to a lanthanide ions (Eu, Pr, Yb are mostly used). Coordination to the lanthanide complex will result in a change in chemical shifts of the nuclei in the analyte dependent on the distance of said nuclei to the lanthanide, and enantiomers will interact differently with the chiral complex, leading to enantiodiscrimination.[237] Diamagnetic chiral solvating agents work in a similar way, though the induced differences in chemical shifts are usually smaller. Here a chiral additive is added to the solvent, which form complexes with the analyte and will result in a change in chemical shifts dependent on the enantiomer, due to differences in interactions. The chiral solvating agents rely heavily on the magnetic anisotropy induced by aromatic groups in the structure of the additives.[237,238]

Also chiral HPLC columns, single crystal X-ray crystallography and vibrational or electronic circular dichroism may be used to discern enantiomers.[232]

## 7.2 Synthesis of media

To widen the pool of enantiodiscriminating alignment media, the focus was decided to be on stretched polymers, since quite good enantiodiscriminative alignment media are available in the LC pool, and due to the inherent problems of strong alignment of LCs. It was thought useful if the enantiodiscrimination of more general compounds than charged amines would be possible utilizing

stretched polymers, and it was proposed to accomplish this by combining the conclusions reported for the polyacetylenes and chiral APhES polymers.[192,210]

The chiral media synthesized were all based on a simple strategy; a chiral building block was attached to a double bond and polymerized with a cross-linker, as illustrated in Figure 7.2. Thus the chiral center is situated in the side-chain of the polymer, while the backbone must be considered a racemic mixture of stereocenters. It is realized that a set chiral center on a side-chain may influence the chirality of the stereocenter set during polymerization, but the tacticity of the polymers was not investigated further.[239–243] The differentiation of enantiomers is proposed to be based on differences in interaction to the chiral centers of the static polymer, leading to a different average orientation in space between enantiomers in the aligned samples.

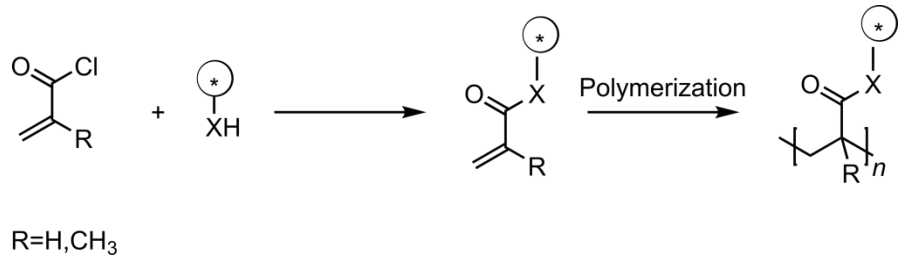



**Figure 7.2. Reaction scheme of the polymerization of chiral monomers (cross-linker not shown).**

The attachment of the double bond to the chiral molecule was achieved by reaction of a secondary alcohol or amine and an acid chloride. This was followed by polymerization in presence of a cross-linker and possibly a co-monomer, using azobisisobutyronitrile (AIBN) as the initiator, outlined in Figure 7.3.

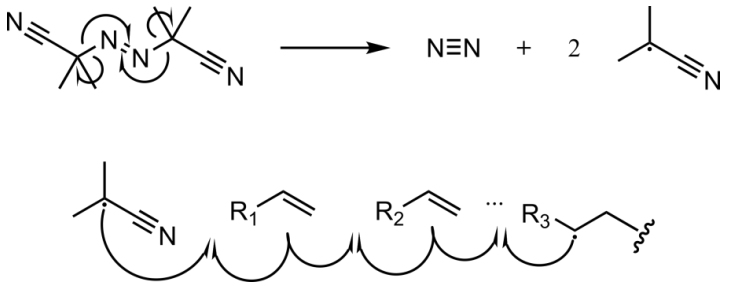



**Figure 7.3. Polymerization reaction using AIBN as an initiator, followed by propagation prolonging the chain until the polymerization is terminated from reaction to a second radical. The initial reaction is induced by heating.**

The chiral molecules used are depicted in Figure 7.4. The molecules were chosen to resemble previously successfully employed monomers from LC and SP alike.[189,192,210] It was important that the dispersion of resonances from the polymer was as small as possible, as residual signals were present in the spectra upon alignment. This limited the amount of possible chiral building blocks to simple structures, or structures with many overlapping shifts, e.g. from long homogenous chains. The importance of this is illustrated in Figure 7.8 page 124, where a spectrum of an aligned sample of strychnine in chloroform is shown. The chosen monomers displayed resonances in two specific regions of NMR spectra: one in the aromatic region and one in the aliphatic region.

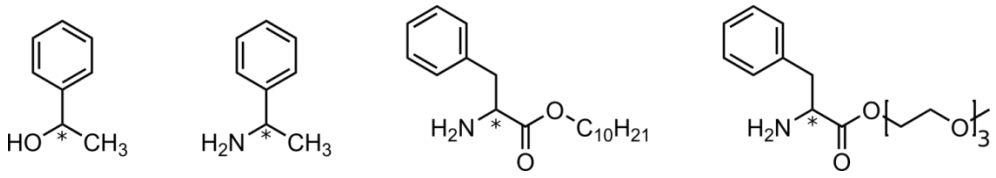


**Figure 7.4. The structures of the chiral building blocks used in the work of this thesis. From left: 1-phenyl ethanol (PhEtO), 1-phenyl ethylamine (PhEtN), $C_{10}H_{21}$-phenylalanine ($C_{10}$-Phe) and methyl-trietyleneglycol-phenylalanine (TEG-Phe).**

The first two chiral building blocks used, (*S*)-1-phenyl ethanol (PhEtO) and (*R*)-1-phenyl ethylamine (PhEtN), were readily available as pure enantiomers, while the remaining two were based on *L*- or *D*-phenylalanine and dubbed $C_{10}H_{21}$-phenylalanine ($C_{10}$-Phe) and methyl-trietyleneglycol-phenylalanine (TEG-Phe) in the following, synthesized as indicated in Figure 7.5.

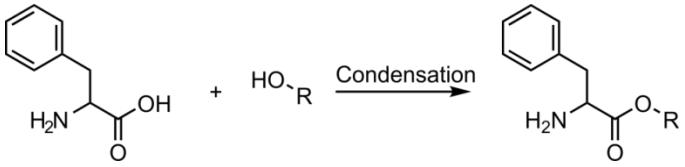



**Figure 7.5. Condensation reaction of phenylalanine under acidic conditions.**

For more on the synthetic condition see the experimental section.

## 7.3 Results

The swelling properties of the resulting polymers were evaluated similarly to the achiral polymers described in Section 6.1. In all cases, in line with available literature and previous findings for the PDMAA polymer, the radial swelling of a thin polymer was superior to vertical swelling of a short polymer, see Figure 5.6 for an illustration.[194,210] This is in part contributed to the fact that the thicker polymer sticks used for vertical swelling formed a plug when solvent was added, which hindered further vertical swelling. Thus only very small RDCs were observed. The nature of the study, comparing enantiomers, also required a high degree of reproducibility, where radial swelling was determined to be inherently superior due to more control of the swelling. Experiments showed that it was much easier to cut the right length of a polymer stick and swell this under identical restriction, than hoping that the polymers would stop swelling at identical lengths. One could of course restrict vertical swelling as well, had plugs not been formed upon the addition of solvent. Another argument in favor of using the radial swelling approach was the time needed for swelling. While vertical swelling may take weeks to months for certain polymers,[191,194] radial swelling was visibly finished and experimentally stable, in regards to the extracted RDCs, in a matter of 5-10 hours in chloroform and 1-2 days in DMSO. No to minor $^2$H-splitting of the solvent was observed for the polymers in all solvents, in contrast to the native achiral polymers, which for the most part was PDMAA (see Section 6.1).

Interfering differences in observed RDCs were determined for identical compounds aligned in identical polymers, which were contributed to random errors as well as small differences in the length of the utilized polymer sticks. This was corrected by choosing a reference RDC across all samples and comparing the relative differences in the observed RDCs, and led to better reproducibility. As the difference in RDCs between enantiomers is proposed to be their relative

orientation compared to the magnetic field, and scaling all RDCs is equal to altering the alignment strength $D_a$, using the relative RDCs will lead to the right conclusion, while eliminating errors from e.g. different polymer lengths. The reference RDC was chosen as a large RDC with a low variance between all data sets, regardless of stereochemistry. The comparison was made utilizing a setup as explained in Figure 7.6.

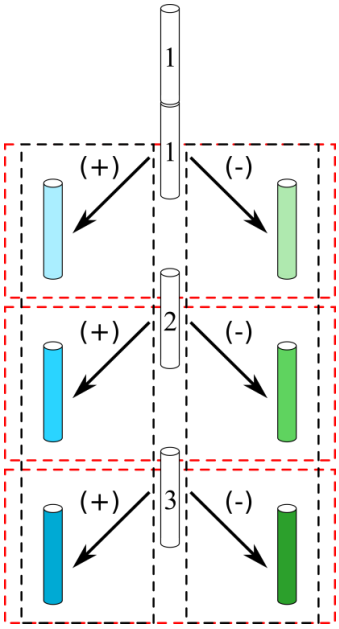



**Figure 7.6. The evaluation of enantiodiscrimination of the alignment media was made by comparing enantiomers, (+) or (-), aligned in polymers from a single polymer stick (red dotted boxes). This was repeated using multiple polymer sticks (2,3…), so that the variation for identical compounds between polymers was also evaluated (black dotted box).**

**This gives rise to two comparisons which are made and depicted in the following sections:**

1. **The averages and standard deviations of RDCs from each black box are compared.**

2. **The RDCs of the red boxes are compared as averages and standard deviations of $RDC_{(+)}$ - $RDC_{(-)}$ for each red box. This is equal to the difference in alignment, disregarding any possible differences in the alignment between polymer sticks.**

## 7.4 PhEtO, PhEtN and $C_{10}$-Phe polymers

It was realized that no enantiodiscrimination could be attributed to the polymers made from (*S*)-1-phenyl ethanol, (*R*)-1-phenyl ethylamine and *L*-$C_{10}H_{21}$-phenylalanine or at least the enantiodiscrimination was below the experimental error and differences between identical compounds. Thus the results will only be shortly presented.

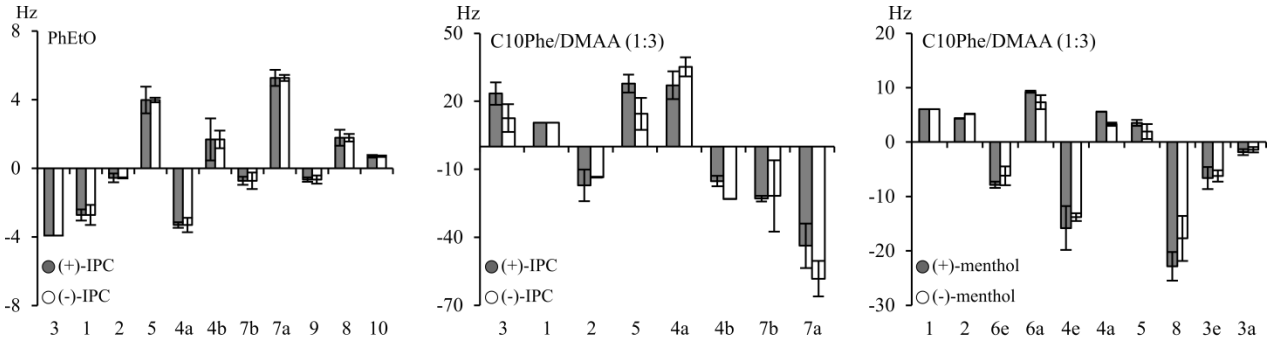



**Figure 7.7. Comparison of RDCs from enantiomers in different chiral media, as given in the plots. Error bars indicate that a mean is used with standard deviation of three datasets for each enantiomer.**

It was possible to synthesize a polymer made entirely of (*R*)-PhEtO, as the monomer was a liquid and solubility was not an issue. The differentiation of enantiomers was not observed though, as illustrated using the enantiomers of IPC in Figure 7.7. The variance between identical compounds

was 0.72 Hz and 0.47 Hz between enantiomers, and it was concluded that enantiodiscrimination was not achieved.

The same trend was observed for the (*R*)-PhEtN polymer, and since it was structurally alike to the (*R*)-PhEtO polymer, while having some solubility difficulties during synthesis, further evaluation was not made. The (*R*)-PhEtN polymer swelled in chloroform, DMSO-$d_6$ and methanol-$d_4$, giving rise to quite different size and sign of RDCs between the apolar and polar solvents.

The (*L*)-$C_{10}$-Phe polymer may differentiate enantiomers, but reproducibility was difficult due to poor solubility of the monomer during synthesis, which resulted in brittle polymers and ultimately large differences in the observed RDCs. A similar LC to the (*L*)-$C_{10}$-Phe polymer has been reported to work well, even at temperatures where no backbone order was supposedly present, and it was surprising that such a low, if any, differentiation was observed.[192] The low solubility of the monomer may be the problem, as the ratio of chiral to achiral monomer might have been too low to achieve observable differentiation.

The solubility problem of the (*L*)-$C_{10}$-Phe polymer was alleviated by using methyl-tri-ethylene glycol as the long ester chain and more of this monomer could be incorporated into the polymer. TEG-phenylalanine based polymers showed minor enantiodiscriminative properties and the results are included below.

## 7.5 The TEG-Phe polymer

The (*L*)-TEG-Phe polymer exhibited much better solubility properties compared to the (*L*)-$C_{10}$-Phe polymer, while retaining approximately the same size of the ester group (though the structure of the chain in solution may vary). This polymer became the focus of the study, and a more in depth analysis of the alignment properties is included. The polymer swelled in chloroform, DMSO-$d_6$ and methanol-$d_4$ giving rise to RDCs in the same general magnitude, but only chloroform and DMSO-$d_6$ was investigated in depth. Similar to the PhEtN polymer, a big difference in extracted relative RDCs was observed between apolar and polar solvents, leading to a proposed difference in alignment, *vide infra*.

The polymer displayed residual resonances grouped in an aromatic (~7-7.5 ppm/120-125 ppm) and a TEG part (~3-4.5 ppm/60-75 ppm) with the latter leading to the most intense signals, illustrated in Figure 7.8. The backbone-, α- and β-resonances were much less pronounced in the spectra, possibly due to less flexibility leading to a shorter $T_2$. The observed analyte resonances were broadened, which may be problematic, due to an increased uncertainty in the extracted coupling constants in the F2 dimension. The determination of the coupling constants was performed by displacement of 1D slices, where the peaks are overlaid to access the values, and the broad peaks should thus not be that problematic as long as they are symmetric.

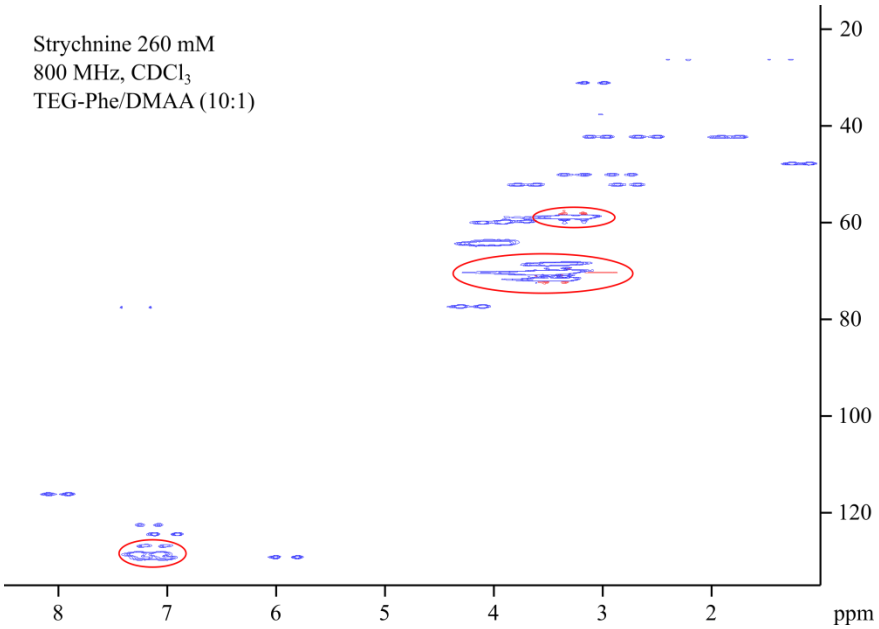



**Figure 7.8. Spectrum of strychnine in $CDCl_3$ aligned in 3 mm TEG-Phe/DMAA polymer stick. Residual polymer signals are highlighted. The only indeterminable strychnine resonance due to spectral overlap is C3-H3 in the aromatic region. The polymer resonances are well contained in small regions of the spectra. (800 MHz, 16 scans, rd=1.5 s, 4k×256 zero-filled to 16k×1k)**

Alignment and differentiation of the enantiomers of the small molecules IPC and menthol was pursued, and a single enantiomer of strychnine was also investigated, using polymers of opposite side-chain chirality.

### 7.5.1 Menthol

Menthol is a molecule with a rigid chair conformation due to three substituents in equatorial positions. A problem in using menthol in RDC comparisons, using only $^1D_{CH}$-coupling constants, is that only four non-parallel internuclear vectors are present, unless the isopropyl orientation is considered known, see Figure 7.9. Consequently, SVD fittings were only illustratively performed from the data. For more details on SVD, the reader is referred to Section 8.1.

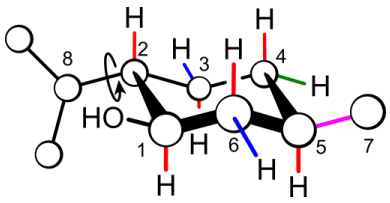



**Figure 7.9. The structure of menthol. Equally colored CH-bonds indicate that the bonds are parallel.**

The general trends of the extracted RDCs could be investigated though, and it was observed that the equatorial protons of the enantiomers had different signs in DMSO. This is a clear sign of enantiodiscrimination as the angle to the magnetic field must differ, independent of alignment strength. A similar trend of differentiation of the equatorial protons, although less pronounced, was determined from alignment in chloroform. Though the structure may not be fixed in space by SVD, the data suggested that the structure was aligned differently between the two solvents, from the sign of the RDCs of the axial protons. Reproducing the RDC results was more difficult in chloroform compared to DMSO, which may be contributed to rapid swelling in the solvent. This correlates well to an observation of the RDCs being markedly different if the insert in the tube had visibly moved.

All samples with a displaced insert were discarded, but small displacement, unnoticeable to the human eye cannot be dismissed.

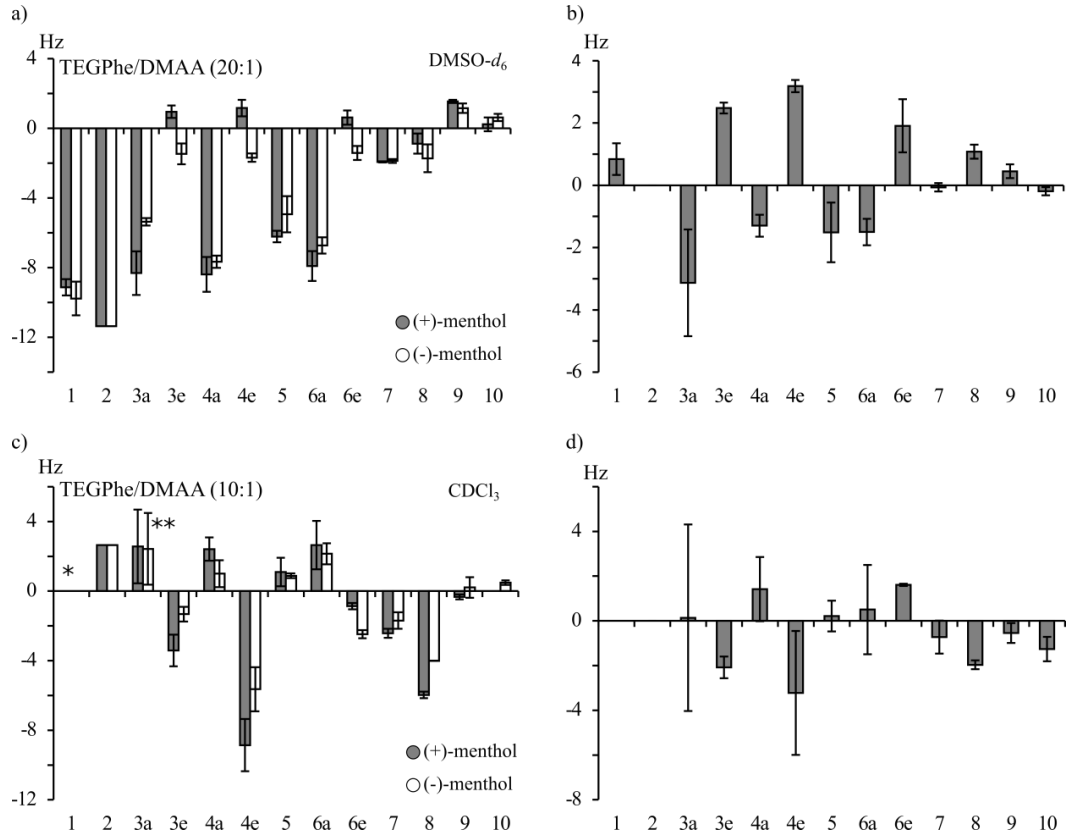



**Figure 7.10. Comparison of RDCs from enantiomers of menthol in (*L*)-TEG-Phe/DMAA. Solvent is DMSO-$d_6$ in a) and b), and chloroform in c) and d). In DMSO three datasets were used for (+)-IPC and two for (-)-IPC. In $CDCl_3$ two datasets were used for each. b) and d) displays values of (-)-menthol subtracted from (+)-menthol in rods cut from identical polymer sticks. A mean is displayed with standard deviation of three datasets for each enantiomer.*Omitted due to overlap with polymer resonances. **Overlaps with CH-7.**

The data was fitted using SVD, though the fit is underdetermined and the structure is placed due to minor difference in bond lengths and angles for parallel vectors in the chair-conformation. Thus experimental errors will have a huge effect. The difference in the experimental RDCs, defined as the generalized cosine β angle (GCB), between identical stereoisomers was 0.996±0.001 or 4.8±0.8° and 0.943±0.007 or 19.3±1.3° between enantiomers, illustrated in Figure 7.11 by 3D structures of (+)-menthol turned into the alignment frames. The definition of GCB is given in Section 5.1.4 page 109.

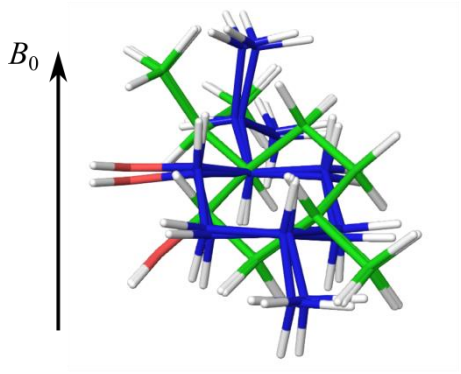



**Figure 7.11. Comparison of the alignment from RDCs of (+)- and (-)-menthol in DMSO-$d_6$ aligned in (*L*)-TEG-Phe/DMAA. Both datasets were applied to the structure of (+)-menthol, green are (-)-menthol and blue (+)-menthol datasets.**

In conclusion, menthol was differentiated by alignment in the (*L*)-TEG-Phe polymer, especially apparent in DMSO, while the exact difference in alignment frame was not accurately determined.

### 7.5.2 IPC

The small molecule IPC was also investigated in DMSO and chloroform. Similar comments to those of menthol on the reproducibility between solvents may be made, and only the data from DMSO are presented.

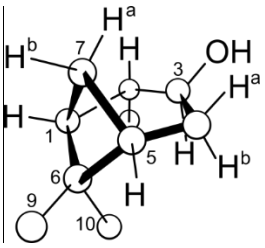



**Figure 7.12. Structure of (+)-IPC.**

The RDCs of aligned enantiomers in DMSO are compared in Figure 7.13, with the dataset including three repetition for each enantiomer. The differences in observed RDCs were small, and again the differences in the used polymer sticks were eliminated by subtraction the RDCs of (-)-IPC from those of (+)-IPC, aligned in sticks cut from the same polymer. A pattern was identified, though the deviations were substantial compared to the observed differences. The GCB between enantiomers was determined to be 0.987±0.005, compared to 0.997±0.002 for identical compounds. This is equal to a difference in angle between alignment frames of 8.9±2.0° between enantiomers compared to 4.4±1.6° for identical compounds. It is thus not a significant differentiation, and it may be argued that the differences are too low to support enantiodiscrimination of IPC due to comparable differences between enantiomers and the reproducibility of the data. Compared to the GCB of PBLG, which have been reported as 0.991, the presented polymer seems to be a slightly better enantiodiscriminator, but exhibit less reproducibility, limiting the credibility.[189,192] The alignment frames of the enantiomers are compared in Figure 7.14.

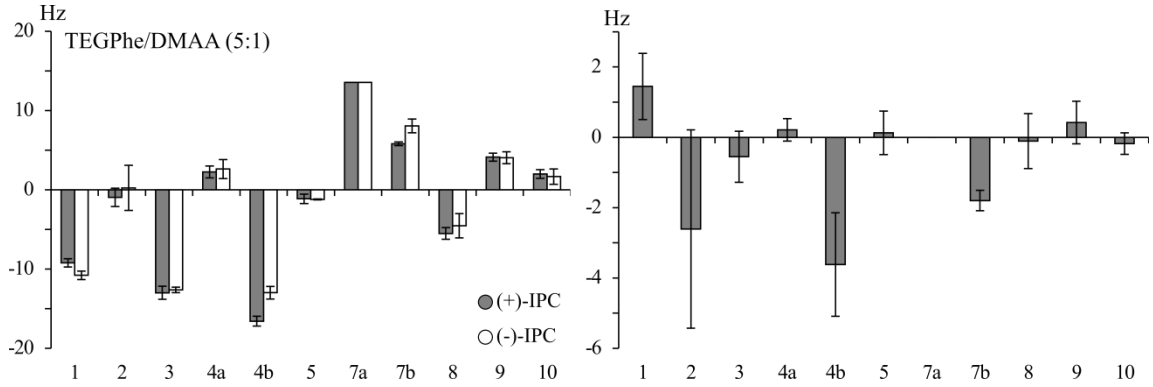



**Figure 7.13 Comparison of RDCs from enantiomers of IPC in (*L*)-TEG-Phe/DMAA (left). Solvent is DMSO-$d_6$. Right: Values of (-)-IPC subtracted from (+)-IPC in rods cut from identical polymer sticks. A mean is used with standard deviation of three datasets for each enantiomer.**

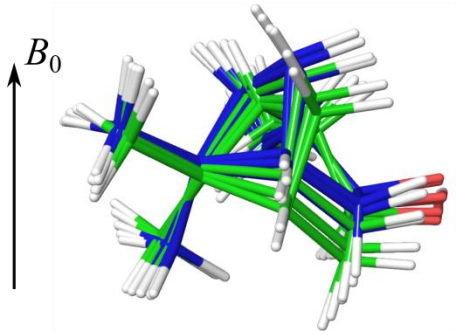



**Figure 7.14. Comparison of the alignment from RDCs of (+)- and (-)-IPC in $CDCl_3$ aligned in (*L*)-TEG-Phe/DMAA. Both datasets were applied to the structure of (+)-IPC, green are (-)-IPC and blue (+)-IPC datasets. The enantiomers are grouped in two, slightly overlapping groups.**

### 7.5.3 Enantiomeric polymers

It is possible to synthesize monomers starting from both enantiomers of phenylalanine and it was thus possible to check whether e.g. strychnine, where only one enantiomer is readily available, could be differentiated by the polymers. Comparing RDCs from the *D*-polymer for (+/-)-menthol to data of the enantiomers from the *L*-polymer, indeed opposite RDCs were obtained, shown in Figure 7.15, where RDCs of (-)-menthol in *D*-TEG-Phe are similar to those of (+)-menthol in *L*-TEG-Phe and vice versa. For the IPC enantiomers the result was less convincing, as seen in Table 7.1. An increase in the GCB of IPC enantiomers in different media was observed compared to identical compounds in identical media, especially for (+)-IPC in *L*-TEG-Phe and (-)IPC in *D*-TEG-Phe, where the difference nears the difference between the enantiomers in both polymers. While this translates to a difference in orientation between the enantiomeric polymers, identical trends in RDCs were observed. Again it would seem that the enantiomers of IPC were not significantly differentiated. This also shows that conclusions based on RDCs obtained between enantiomeric polymers needed to be made carefully.

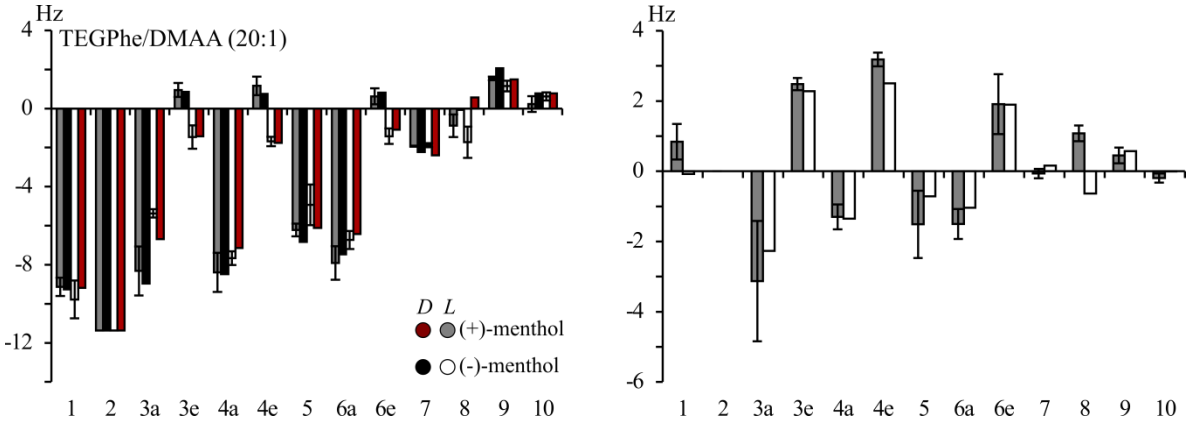



**Figure 7.15. Left: Comparison of RDCs from enantiomers of menthol in TEG-Phe/DMAA. Solvent is DMSO-$d_6$. Grey (+) and white (-) are the enantiomers in *L*-Phe based polymers, while black (-) and red (+) are enantiomers in *D*-Phe based polymers. Right: Values of (-)-menthol subtracted from (+)-menthol in rods cut from identical polymer sticks. Grey is *L*-Phe and white is *D*-Phe based polymer. A single dataset for the enantiomers in D-Phe based polymers was used.**

**Table 7.1. Comparison of the alignment tensors of enantiomers of IPC from *L* and *D*-variant of the TEG-Phe based alignment media. The comparison is performed by the GCB. Comparisons which should lead to equal alignment are marked in blue, and different alignments in red, if enantiodiscrimination is achieved. Dataset of *L*-polymers include three sets of RDCs for each enantiomer, for the *D*-polymer two are included.**

| | | ***L*-Phe** | | ***D*-Phe** | |
|---|---|---|---|---|---|
| | | (+)-IPC | (-)-IPC | (+)-IPC | (-)-IPC |
| *L*-Phe | (+)-IPC | 5.0±1.7° | 8.9±2.0° | 9.6±3.1° | 8.2±2.0° |
| | (-)-IPC | | 1.7±1.5° | 4.1±1.3° | 9.9±2.2° |
| *D*-Phe | (+)-IPC | | | 6.4° | 9.7±3.0° |
| | (-)-IPC | | | | 4.8° |

### 7.5.4 Strychnine

As the enantiomeric polymers aligned single enantiomers differently, at least in the case of menthol, a theoretical differentiation of enantiomer of strychnine was approached, similar to the work of Schmidt *et al.*[210] Not all one-bond vectors were used for comparison, as the resonance of C3-H3 were overlapped with polymer resonances in the CLIP-HSQC spectra and the RDCs could not be determined. The resonances of C-15 were weak and slightly distorted compared to the other resonances and were omitted (the extracted RDCs were of equal size between all samples). The resonances of C-23 were not separated, in line with experiences from other media (Section 6.2). As a result 15 RDCs were extracted, and a difference was observed between the polymers based on *L*- and *D*-phenyl alanine, as illustrated in Figure 7.16.

The data was based on a triple study, using three different *L*- and *D*-polymer sticks. While it was not possible to compare alignment from equal polymer sticks, as an identical enantiomer of strychnine was used throughout, the alignment was compared based on the succession of obtain data, i.e. polymer $L_1$ is compared to $D_1$, $L_2$ to $D_2$ and $L_3$ to $D_3$, in Figure 7.16. The used polymers sticks were synthesized from the same bulk reaction mixture, but polymerized in different tubes, as to take differences in polymerization into account.

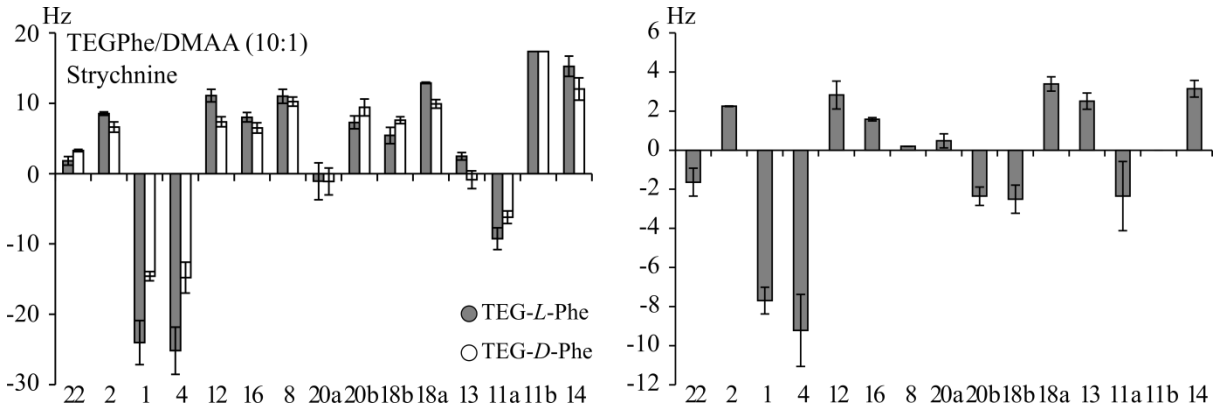



**Figure 7.16. Left: Comparison of RDCs from strychnine in (*L*)-TEG-Phe/DMAA (grey) and (*D*)-TEG-Phe/DMAA (white). Solvent is $CDCl_3$. Right: Values obtained by subtracting polymer $D_1$ from $L_1$, $D_2$ from $L_2$ and $D_3$ to $L_3$. A mean is used with standard deviation of three datasets for each enantiomeric polymer.**

Rather big differences in RDCs between strychnine aligned in the two media were observed, and enantiodiscrimination was, cautiously, considered achieved. Examples of 1D slices through the CLIP-HSQC are found in Figure 7.17.

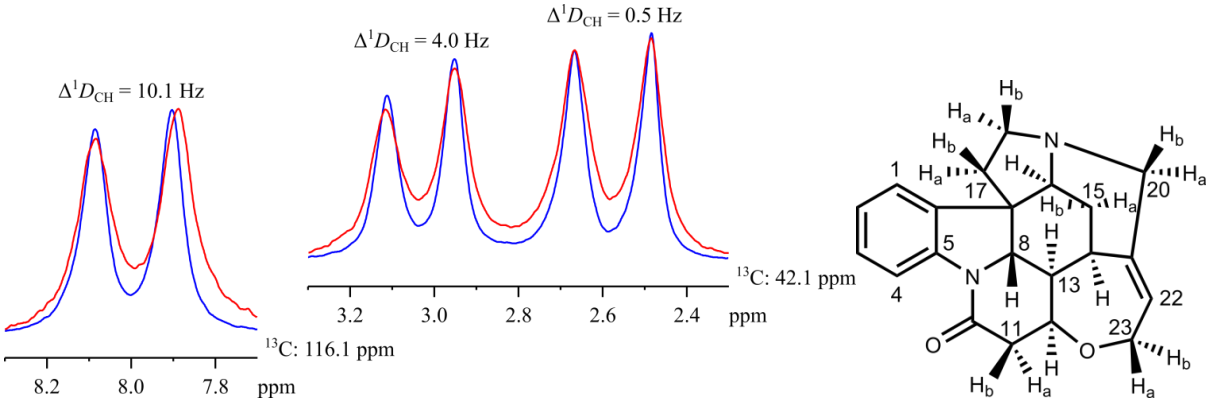



**Figure 7.17. 1D slices from a CLIP-HSQC of strychnine in (*L*)-TEG- Phe/DMAA (blue) and (*D*)-TEG-Phe/DMAA (red) and the differences in the RDCs between the polymers. The total coupling constants were determined by displacement of peaks and comparison to extracted *J*-coupling constants (isotropic spectra not shown).**

However, a problem was identified. While the SVD correlations of the experimental to back-calculated data for IPC enantiomers were generally good, with *Q*-factors between 0.085 and 0.210 (average 0.135), the correlation of the data to a strychnine 3D structure was bad, with *Q*-factors between 0.247 and 0.374 (average 0.314). It was investigated whether the problem could arise from the strychnine data being acquired at 800 MHz, compared to 400 MHz for IPC and menthol. When comparing datasets from 400 MHz to those of 800 MHz, the values varied slightly with an RMSD of 2.0 Hz, but the resulting correlation between experimental data and 3D structure was not improved compared to that of the 800 MHz data. Multiple resonances were less resolved at 400 MHz, the reason for acquiring the dataset at 800 MHz initially, and if these RDCs are omitted in the comparison the RMSD is 1.1 Hz between the two field strengths. The cause of the large difference between experimental and back-calculated data was not identified, and the reproducibility contradicts a conclusion of the difference being a product of erroneous extraction of RDCs. Also known parallel vectors (e.g. C1-H1 and C4-H4) have equal RDCs, which initially indicated that experimental RDCs would fit back-calculated values.

When comparing alignment tensors, which will be less accurate due to the bad correlation between experimental RDCs and the SVD calculations, a difference between the two polymers was identified. The alignment frames from RDCs differ comparable to e.g. IPC in DMSO, with a GCB of 0.990±0.006 or 7.4±2.7° between identical media. Between different media the GCB was calculated to be much lower (resulting in a larger angle): 0.940±0.037 or 19.0±6.2°.

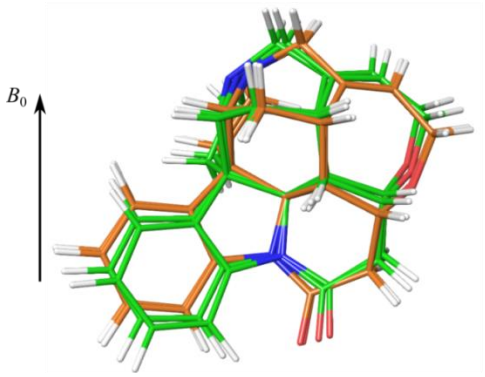



**Figure 7.18. Comparison of the alignment from RDCs of strychnine in (*L*)-TEG-Phe/DMAA (green) and (*D*)-TEG-Phe/DMAA (orange).**

The alignment is markedly different compared to the PDMAA polymer, the pure polymer of the co-polymer used with TEG-Phe, as seen in Figure 7.19. This may not be surprising due to the low amount of DMAA used in polymerization (10 mol% for strychnine). Back-calculation of RDCs in PDMAA placed the structure with the largest surface horizontal compared to the magnetic field, while the TEG-Phe polymer placed the largest surface parallel to $B_0$, perhaps best envisioned by the sign of the RDCs of C1-H1 and C4-H4 in Figure 7.19. Since identical experimental procedures were followed, restricted swelling of 3 mm rods, this could translate to the TEG-chains extending from the surface with the analyte situated between the chains. The situations are illustrated by a cartoon in Figure 7.19.

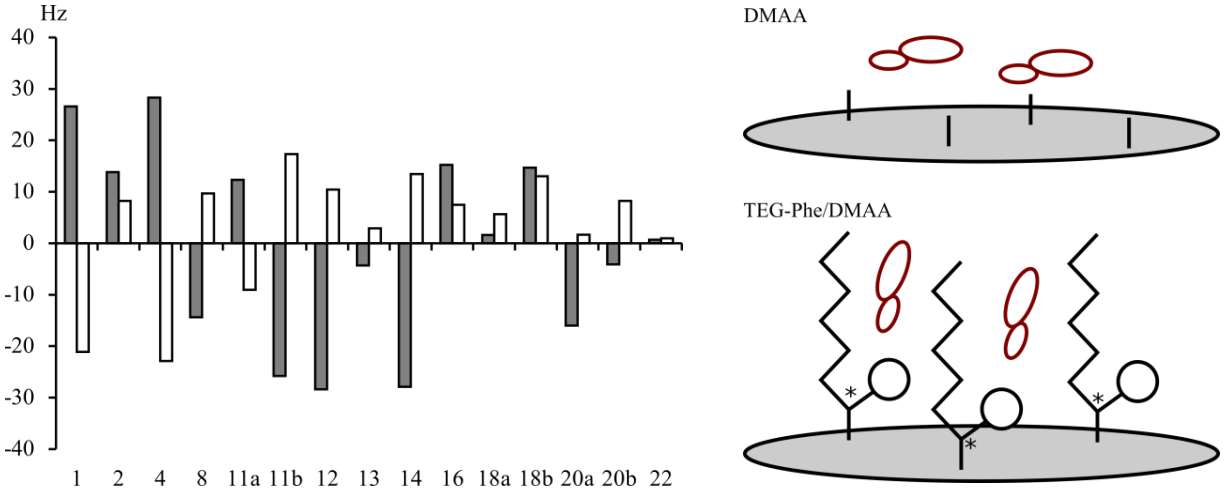



**Figure 7.19. Left: Example of RDCs from strychnine aligned in PDMAA (grey) and (*L*)-TEG-Phe/DMAA (white). Right: A possible scenario of alignment which explains the differences in RDCs. Red figures are analytes, the grey surface is an near-infinite polymer surface and the polymer side-chains and alignment of analytes are indicated.**

When stretched in DMSO, the polymer aligned solutes, such as IPC and menthol, perpendicular to the results in chloroform and more alike to the DMAA polymer (which also swells in DMSO).

## 7.6 Perspectives

For RDCs to impact the small molecule community in regard to enantiodiscrimination a couple of challenges need to be addressed, none of which are solved by the polymers introduced in this thesis. The largest problem is the need for both enantiomers of an analyte in comparison, due to a low difference in alignment between enantiomers. This is a problem in determining an unknown absolute stereochemistry of e.g. a natural product. For this feat to be realized, RDCs need to be more integrated with computational chemistry. The possibility to couple computational chemistry to RDCs and determine the absolute stereochemistry has been disputed, and it is acknowledged that the technology is not yet available to achieve this.[244]

If the difference in alignment is large between enantiomers, as for e.g. shown for IPC aligned in the acetylenes,[189,192] and with further development of the methodology of Luy *et al.* where the alignment was determined including a polymer strand in simulation of strychnine, the possibilities are enticing.[245] If simulations could capture the differences in interactions of enantiomers to a chiral polymer strand, it should be possible to determine the absolute stereochemistry from a single enantiomer. This would require a difference in theoretical GCB of more than the 10-20° determined in this study, to get useful and trustworthy results and to negate errors from simulations and determined RDCs.

It would be interesting to investigate whether the TEG-Phe group, which showed better properties in stretched polymer, would retain those properties if utilized in LCs. This might be in LCs such as the acetylenes,[189,192] where an LC that align molecules in both chloroform and DMSO might be obtained. A proposed synthesis is outlined in Figure 7.20.

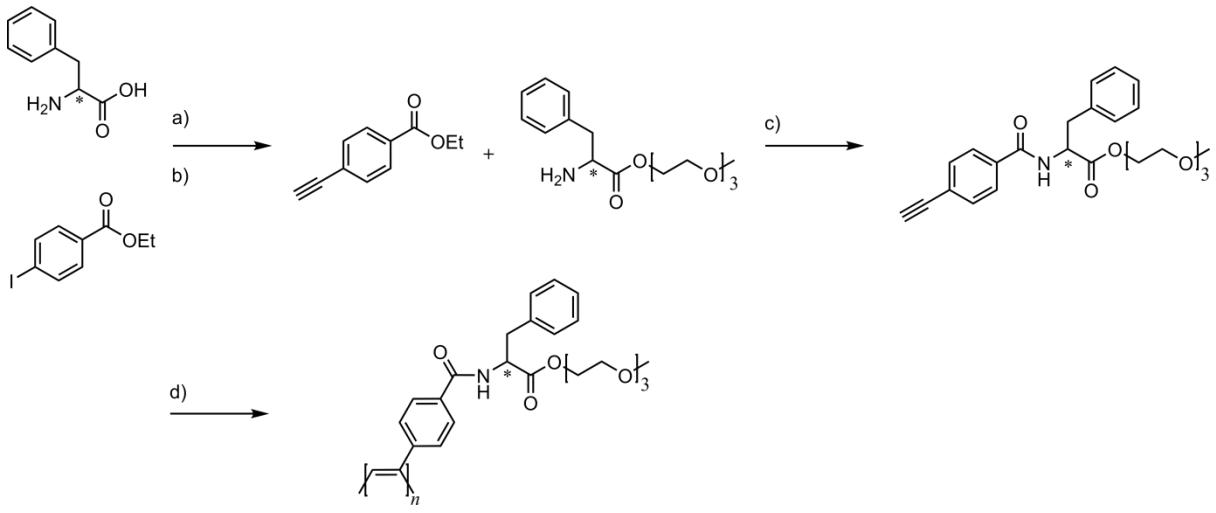



**Figure 7.20. Possible synthetic route to obtain an acetylene-based LC that might work in DMSO and $CDCl_3$, based on the stretched polymer work presented in this thesis. a) see experimental section. b)-d) see work of Krupp and Reggelin.[192]**

Also other LCs might be of interest. Griesinger *et al.* recently published a LC made of graphene oxide with attached polymer brushes made of trifluoroethyl methacrylate, which displayed very interesting properties such as very narrow line widths and effectively no residual polymer signals.[214]

The utilized solvent is DMSO, and due to the swelling properties, and thus solubility, of the TEG-Phe polymer in this solvent, it could easily be envisioned as an enticing prospect of adding chirality to the achiral graphene sheets and possibly obtain enantiodiscrimination.

Based on the assumption that the (*L*)-$C_{10}$-Phe polymer failed to discriminate enantiomers due to a low amount of the chiral monomer incorporated into the polymers, it was not consistently tried to synthesize polymer with a molar fraction of TEG-Phe below 80 %. In retrospect, it might be that a lower degree of the larger chiral monomer would lead to better access to the chiral centers, and thus better differentiation. An optimal molar fraction of chiral monomer for differentiation may thus be lower than the ones tested, which should be investigated further.

## 7.7 Conclusions

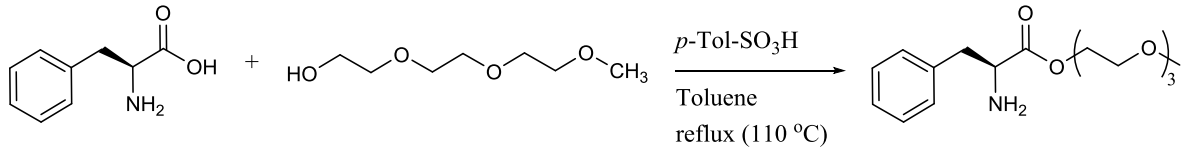
In conclusion a polymer was identified that possess minor enantiodiscriminating properties comparable to PBLG in chloroform and DMSO. In principle a similar differentiation should be possible in methanol as the polymers align compounds in this solvent at an equal magnitude of alignment strength as DMSO, but this was not attempted. The line width might be problematic and further studies are needed to verify the differentiation. While being minor, the differentiation of enantiomers show promise and this is the first uncharged radical-initiated polymer that has been shown to possess enantiodiscriminative properties. As for all alignment media, it is still a requirement to have each enantiomer as a standard for comparison, which renders the technology difficult and slow to use compared to e.g. optical rotation, which will result in similar, and possibly more significant, information regarding absolute chemistry.

## 7.8 Experimental

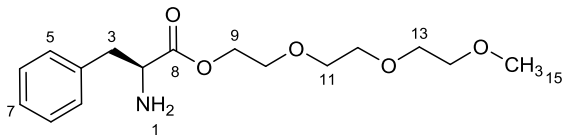
**Synthesis of chiral polymers**

The chemicals used were commercially available and the solvents used were all HPLC grade. Dry solvents were obtained from. All flasks were dried over a Bunsen flame under vacuum before any reactant or solvent was added. All flasks (or NMR tubes) were equipped with a rubber septum, and chemicals were transported by syringe. All commercially bought monomers and crosslinkers (except *N,N*-methylenbisacrylamide which is a solid) were purified prior to the synthesis to remove the polymerization inhibitor by passing the neat liquid through a pipette filled with basic alumina.

**Preparation of (triethylene glycol methyl ether)-L-phenylalaninate**




10 g L-phenylalanine (61 mmol) was suspended in 250 mL toluene, followed by 10.4 mL triethylene glycol methyl ether (67 mmol) and 13.8 g p-toluenesulfonic acid monohydrate (73 mmol). The suspension was fitted to with a Dean and Stark water trap and heated to reflux until the calculated amount of water was collected (2.4 mL). The mixture was concentrated to dryness under reduced pressure to yield a yellowish oil. 250 mL DCM was added and the solution washed with 2x150 mL saturated aqueous sodium carbonate solution and once with 100 mL water, dried over $Na_2SO_4$ and concentrated to yield a yellowish oil. Yield: 15.6 g (75 %).




$[\alpha]_D^{20}$ = +28.1 (c = 0.006, MeOH). MS (ESI): m/z: 312.2 $[M+H]^+$. m/z calcd. for $C_{19}H_{32}NO_2$ $[M+H]^+$: 312.18.

$^1$H NMR ($CDCl_3$, 400 MHz, 298 K): δ = 2.19 (2H(1), m), 2.96 (1H(3b), dd, J=13.5,7.6 Hz), 3.15 (1H(3a), dd, J=13.5,5.3 Hz), 3.40 (3H(15), s), 3.57 (2H(14), m), 3.68 (6H(11-13), m), 3.71 (2H(10), m), 3.83 (1H(2), broad-t, m), 4.30 (2H(9), t, J=4.8 Hz), 7.25 (2H(5), m), 7.26 (1H(7), m), 7.32 (2H(6), m)

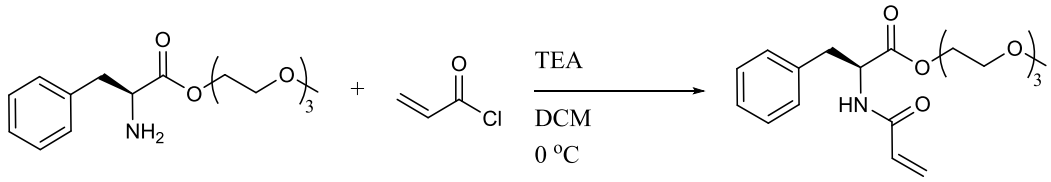
$^{13}$C NMR ($CDCl_3$, 125 MHz, 298 K): δ = 40.7 (C-3b) , 40.7 (C-3a) , 55.6 (C-2) , 59.0 (C-15) , 64.0 (C-9) , 69.1 (C-10) , 70.7 (C-11-13) , 72.0 (C-14) , 127.0 (C-7) , 128.8 (C-6) , 129.4 (C-5) , 137.1 (C-4) , 174.7 (C-8)

**Preparation of (triethylene glycol methyl ether)-D-phenylalaninate**

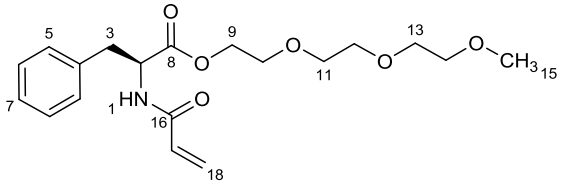
As for (triethylene glycol methyl ether)-L-phenylalaninate described above.

$[\alpha]_D^{20}$ = -17.1 (c = 0.007, MeOH).

**Preparation of (triethylene glycol methyl ether)-acryloyl-L-phenylalaninate**




4.5 g (Triethylene glycol methyl ether)-L-phenylalaninate (15 mmol) and 2.4 mL TEA (18 mmol) was added to 200 mL dry DCM in a flame dried round-bottomed flask under argon and the solution was cooled to 0 °C. 1.3 mL acryloyl chloride (16 mmol) was added drop wise over approximately 30 min. The solution was kept at 0 °C for 2 h and was then left at rt for an additional 2 h. The solution was washed with 2x100 mL 0.5 M HCl and once with 50 mL water, dried over $Na_2SO_4$ and concentrated to a yellowish oil. Freeze dried to yield a white solid. Yield: 4.5 g (84 %).



$[\alpha]_D^{20}$ = +33.2 (c = 0.004, MeOH). MS (ESI): m/z: 366.2 $[M+H]^+$. m/z calcd. for $C_{19}H_{32}NO_2$ $[M+H]^+$: 366.19.

$^1$H NMR ($CDCl_3$, 400 MHz, 298 K): δ = 3.19 (2H(3), m), 3.35 (3H(15), s), 3.53 (2H(14), m), 3.64 (6H(11-13), m), 3.68 (2H(10), m), 4.28 (2H(9), m), 4.99 (1H(2), dt, J=7.7,5.6 Hz), 5.65 (1H(18b), dd, *J*=10.3,1.3 Hz), 6.09 (1H(17), dd, J=17.0,10.3 Hz), 6.15 (1H(1), d, J=7.8 Hz), 6.28 (1H(18a), dd, J=16.9,1.4 Hz), 7.12 (2H(5), dd, J=7.8,1.2 Hz), 7.23 (1H(7), m), 7.26 (2H(6), m)

$^{13}$C NMR ($CDCl_3$, 125 MHz, 298 K): δ = 37.7 (C-3), 53.1 (C-2), 58.9 (C-15), 64.6 (C-9), 68.8 (C-10), 70.6 (C-11-13), 71.9 (C-14), 127.1 (C-7), 127.2 (C-18), 128.7 (C-6), 129.3 (C-5), 130.5 (C-17), 135.7 (C-4), 164.8 (C-16), 171.5 (C-8)

**Preparation of (triethylene glycol methyl ether)-acryloyl-D-phenylalaninate**

As for (triethylene glycol methyl ether)-acryloyl-L-phenylalaninate described above.

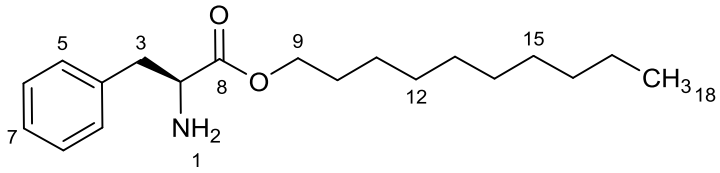
$[\alpha]_D^{20}$ = -24.7 (c = 0.003, MeOH).

**Preparation of poly-(triethylene glycol methyl ether)-acryloyl-L-phenylalaninate/*N*,*N*-dimethyl acrylamide (p-TEGPhe/DMAA)**

0.533 g of (Triethylene glycol methyl ether)-acryloyl-L-phenylalaninate (1.5 mmol) was dissolved in 0.7 mL $CDCl_3$. 0.2 mL *N*,*N*-dimethylacrylamide (0.1 mmol) and 3.4 mg *N*,*N*-methylenbisacrylamide (0.01 mmol) was added. 1.5 mg AIBN (0.001 mmol) was added and $N_2$ bubbled through the solution for 5 minutes to remove $O_2$. The solution was transferred to 3 mm (or 5 mm) NMR tubes. Air was removed under vacuum and $N_2$ backfilled into the tubes three times and the tubes were left at 60 °C overnight (approx. 14-16 h) to polymerize.

**Preparation of decyl-L-phenylalaninate**

10 g L-phenylalanine (61 mmol) was suspended in 250 mL toluene, followed by 12.7 mL decanol (67 mmol) and 13.8 g p-toluenesulfonic acid monohydrate (73 mmol). The suspension was fitted to with a Dean and Stark water trap and heated to reflux until the calculated amount of water was collected (2.3 mL). The mixture was concentrated to dryness under reduced pressure to yield a yellowish oil. 250 mL DCM was added and the solution washed with 2x150 mL saturated aqueous sodium carbonate solution and once with 100 mL water, dried over $Na_2SO_4$ and concentrated to yellowish oil. Re-crystalized from EtOH twice. Yield: 10.0 g (HCl salt) (48.3 %).



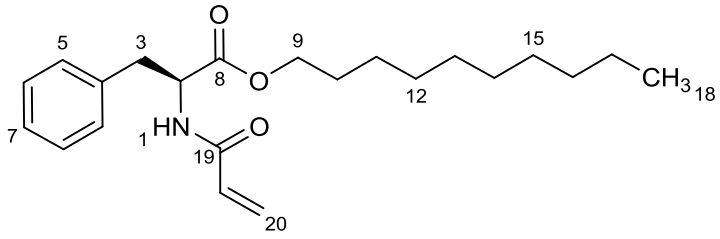

$[\alpha]_D^{20}$ = +14.7 (c = 0.008, $CHCl_3$). MS (ESI): m/z: 306.2 $[M+H]^+$. m/z calcd. for $C_{19}H_{32}NO_2$ $[M+H]^+$: 306.24.

$^1$H NMR ($CDCl_3$, 400 MHz, 298 K): δ = 0.88 (3H(18), t, *J*=7.0), 1.19-1.35 (14H(11-17), m), 1.59 (2H(10), m), 2.53 (2H(1), broad s), 2.94 (1H(3a), dd, *J*=13.7,7.7), 3.11 (1H(3b), dd, *J*=13.6,5.7), 3.80 (1H(2), dd, *J*=7.3,5.6), 4.09 (2H(9), t, J=6.7 Hz), 7.17-7.33 (5H(5-7), m)

$^{13}$C NMR ($CDCl_3$, 125 MHz, 298 K): δ = 14.2 (C-18), 22.4 (C-17), 25.7, 29.3, 31.8 (C-11-16), 28.4 (C-10), 40.3 (C-3), 55.4 (C-2) , 65.2 (C-9) , 126.8, 128.5, 129.2 (C-5-7), 136.6 (C-4), 174.1 (C-8)

**Preparation of decyl-acryloyl-L-phenylalaninate**

1.5 g decyl-L-phenylalaninate (4.9 mmol) and 0.82 mL TEA (5.9 mmol) was added to 200 mL dry DCM in a flame dried round-bottomed flask under argon and the solution was cooled to 0 °C. 0.44 mL acryloyl chloride (5.4 mmol) was added drop wise over approximately 30 min. The solution was kept at 0 °C for 2 h and was then left at rt for an additional 2 h. The solution was washed with 2x100 mL 0.5 M HCl and once with 50 mL water, dried over $Na_2SO_4$ and concentrated to white solid. Yield: 1.77 g (95.2 %).




$[\alpha]_D^{20}$ = +17.2 (c = 0.005, $CHCl_3$). MS (ESI): m/z: 360.3 $[M+H]^+$. m/z calcd for $C_{22}H_{34}NO_3$ $[M+H]^+$: 360.25.

$^1$H NMR ($CDCl_3$, 400 MHz, 298 K): δ = 0.89 (3H(18), t, *J*=7.1), 1.20-1.35 (14H(11-17), m), 1.59 (2H(10), m), 3.17 (2H(3), m), 4.11 (2H(9), m), 4.96 (1H(2), dt, J=7.7,5.8 Hz), 5.68 (1H(20a), dd, J=10.3,1.4 Hz), 6.02 (1H(1), d, J=7.4 Hz), 6.09 (1H(19), dd, J=17.0,10.3 Hz), 6.29 (1H(20b), dd, J=16.9,1.3 Hz), 7.09 (2H(5), m), 7.22-7.31 (3H(6-7), m)

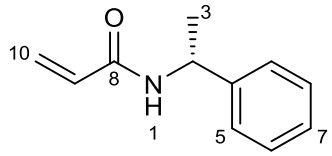
$^{13}$C NMR ($CDCl_3$, 125 MHz, 298 K): δ = 13.8 (C-18), 22.4 (C-17), 25.7, 29.3, 31.8 (C-11-16), 28.3 (C-10), 37.9 (C-3), 52.9 (C-2) , 65.8 (C-9) , 127.1 (C-20), 127.2, 128.5 (C-6-7), 129.2 (C-5), 130.3 (C-19), 135.7 (C-4), 164.6 (C-16), 171.8 (C-8)

**Preparation of poly-decyl-acryloyl-L-phenylalaninate/*N*,*N*-dimethyl acrylamide (p-$C_{10}$Phe/DMAA)**

1.93 g of decyl-acryloyl-L-phenylalaninate (5.4 mmol) was dissolved in 1.5 mL $CDCl_3$. 1.5 mL *N*,*N*-dimethylacrylamide (14.6 mmol) and 13.0 mg *N*,*N*-methylenbisacrylamide (0.08 mmol) was added. 2.9 mg AIBN (0.002 mmol) was added and $N_2$ bubbled through the solution for 5 minutes to remove $O_2$. The solution was transferred to 3 mm (or 5 mm) NMR tubes. Air was removed under vacuum and $N_2$ backfilled into the tubes three times and the tubes were left at 60 °C overnight (approx. 14-16 h) to polymerize.

**Preparation of (*R*)-*N*-(1-phenylethyl)-acrylamide**

2.86 g (*R*)-1-phenylethylamine (23.5 mmol) and 3.9 mL TEA (28.3 mmol) was added to 200 mL dry DCM in a flame dried round-bottomed flask under argon and the solution was cooled to 0 °C. 2.11 mL acryloyl chloride (25.9 mmol) was added drop wise over approximately 30 min. The solution was kept at 0 °C for 2 h and was then left at rt for an additional 2 h. The solution was washed with 2x100 mL 0.5 M HCl and once with 50 mL water, dried over $Na_2SO_4$ and concentrated *in vacuo* to yield a white solid. Yield: 4.03 g (97.6 %).




$[\alpha]_D^{20}$ = +22.2 (c = 0.009, MeOH). MS (ESI): m/z: 176.1 $[M+H]^+$. m/z calcd for $C_{11}H_{14}NO$ $[M+H]^+$: 176.11.

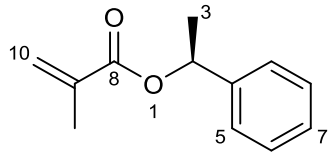
$^1$H NMR (DMSO-$d_6$, 400 MHz, 298 K): δ = 1.38 (3H(3), d, *J*=7.3), 5.0 (1H(2), p, *J*=7.5), 5.60 (1H(10a), dd, *J*=10.1,2.2), 6.08 (1H(10b), dd, *J*=17.1,2.2), 6.29 (1H(9), dd, *J*=17.1,10.2), 7.21 (1H, m), 7.29-7.39 (4H(5-6), m), 7.76 (1H(1), d, J=7.8 Hz)

$^{13}$C NMR ($CDCl_3$, 125 MHz, 298 K): δ = 22.7 (C-3), 48.2 (C-2), 125.7 (C-10), 126.4 (C-5), 127.2 (C-7), 128.7 (C-6), 132.2 (C-9), 144.9 (C-4), 163.9 (C-8)

**Preparation of poly-(*R*)-*N*-(1-phenylethyl)-acrylamide/*N*,*N*-dimethyl acrylamide (p-PhEtN/DMAA)**

0.23 g of (*R*)-*N*-(1-phenylethyl)-acrylamide (1.3 mmol) was dissolved in 0.2 mL acetone-$d_6$. 0.25 mL *N*,*N*-dimethylacrylamide (2.4 mmol) and 2.7 mg *N*,*N*-methylenbisacrylamide (0.02 mmol) was added. 2.8 mg AIBN (0.002 mmol) was added and $N_2$ bubbled through the solution for 5 minutes to remove $O_2$. The solution was transferred to 3 mm (or 5 mm) NMR tubes. Air was removed under vacuum and $N_2$ backfilled into the tubes three times and the tubes were left at 60 °C overnight (approx. 14-16 h) to polymerize.

**Preparation of (*S*)-1-phenylethyl methacrylate**
2.0 mL (*S*)-1-phenylethanol (16.5 mmol) and 2.8 mL TEA (19.9 mmol) was added to 200 mL dry DCM in a flame dried round-bottomed flask under argon and the solution was cooled to 0 °C. 1.8 mL methacryloyl chloride (18.2 mmol) was added drop wise over approximately 30 min. The solution was kept at 0 °C for 2 h and was then left at rt for an additional 2 h. The solution was washed with 2x100 mL 0.5 M HCl and once with 50 mL water, dried over $Na_2SO_4$ and concentrated *in vacuo* to yield a reddish oil. Yield: 2.54 g (80.5 %).




$[\alpha]_D^{20}$ = -29.4 (c = 0.008, MeOH). MS (ESI): m/z: 191.1 $[M+H]^+$ m/z calcd for $C_{12}H_{15}O_2$ $[M+H]^+$: 191.11.
$^1$H NMR ($CDCl_3$, 400 MHz, 298 K): δ = 1.50 (3H(3), d, *J*=6.7), 1.89 (3H(11), broad s), 5.49 (1H(10a), p, *J*=1.7), 5.87 (1H(2), q, *J*=6.6), 6.08 (1H(10b), dq, *J*=1.6,0.8), 6.29 (1H(9), dd, *J*=17.1,10.2), 7.21 (1H, m), 7.24-7.32 (4H(5-6), m)
$^{13}$C NMR ($CDCl_3$, 125 MHz, 298 K): δ = 18.3 (C-11), 22.3 (C-3), 72.4 (C-2), 125.4 (C-10), 126.0 (C-5), 127.8 (C-7), 128.5 (C-6), 136.6 (C-9), 141.7 (C-4), 166.6 (C-8)

**Preparation of poly-(*S*)-*O*-(1-phenylethyl)-methacrylate (p-PhEtO)**
0.57 g of (*S*)-1-phenylethyl methacrylate (3.0 mmol) was dissolved in 0.17 mL acetone-$d_6$. 1.4 µL EGDMA (0.007 mmol) was added. 1.3 mg AIBN (0.001 mmol) was added and $N_2$ bubbled through the solution for 5 minutes to remove $O_2$. The solution was transferred to 3 mm (or 5 mm) NMR tubes. Air was removed under vacuum and $N_2$ backfilled into the tubes three times and the tubes were left at 60 °C overnight (approx. 14-16 h) to polymerize.

**Swelling of polymers**
See Section 6.4 regarding polymers from 3 mm NMR tubes. If the inserted plug visibly moved during swelling the sample was excluded from the study of chiral polymers, due to an observed large error between alignments of equal enantiomers when this happened. RDCs could still be extracted though so this is not a problem for achiral polymers.

Also see Section 6.4 for information regarding synthesis of the PDMAA polymer. All other necessary information is available in figures and tables throughout this chapter.

**NMR experiments**
The CLIP-HSQC experiments was used for the determination of one-bond CH coupling constants, using standard setup as described in Appendix A1. The concentration of analytes was ~150 mM for isotropic samples and ~220 mM (strychnine) or ~320 mM (IPC/menthol) for aligned samples.

**Simulations**
Structures were generated by the modelling suite Maestro version 10.2.010 (2015) by Schrödinger for force field calculations,[126] using the program MacroModel version 10.8.[127,128] The MMFFs force

field was used. To generate structures which should cover the conformational space of compounds, a conformational search was performed for each structure by the program MacroModel using energy cutoff of 50 kJ/mol, 100,000 steps and CPRG minimization.[127,129]

Gaussian version 09 revision B.01 was used for DFT calculations including optimizations and NMR calculations.[131] Structures were optimized to a B3LYP/6-31(d) level of theory.

# 8 Tensor free RDC calculations

For the theoretical basis of RDCs, the reader is referred to Chapter 5. The theory behind the actual back-calculations of RDCs from experimental data and 3D structures will be described briefly below to introduce a new method of back-calculation. The introduction will focus entirely on singular value decomposition (SVD), since this is by far the most utilized method in small molecule RDC calculations. To support a possible need for a new back-calculation method, it is important to understand the basics of SVD and the advantages and disadvantages involved when using this method.

## 8.1 Singular value decomposition

Losonczi *et al.* described the mathematical basis of the SVD method for back-calculation of RDCs of macromolecules.[202] Here the principles of SVD are included in equations (8.1) to (8.12).

Tensors in RDC calculations are based on an expanded form of equation (5.2), Section 5.1. Here, the alignment frame is allowed to shift in comparison to the direction of the magnetic field (usually chosen as the z-axis) and now includes an axial ($A_a$) and rhombic ($A_r$) component, see equation (8.1).[197,203,204,246]

$$D_{IS} = -\frac{\hbar\gamma_I\gamma_S\mu_0}{16\pi^2 r_{IS}^3}\left[A_a(3\cos^2\theta_{IS} - 1) + \frac{3}{2}A_r\sin^2\theta_{IS}\cos 2\varphi_{IS}\right] \quad \textbf{(8.1)}$$

Where $\gamma$ is the gyromagnetic ratio, $\hbar$ is the Planck constants divided by $2\pi$, $\mu_0$ is the vacuum permeability constant, $r$ is the cubed distance between the nuclei and $\theta$ is the angle between the internuclear vector and the magnetic field. $\varphi$ is the azimuthal angle that describes the projection of the RDC vector to the xy-plane with respect to the x-axis, as illustrated Figure 8.2, left.

This will lead to an averaging of the internuclear vectors positioned along the x-axis and is needed to unambiguously place the molecules in the alignment frame. Four possible alignments are always present due to equal theoretical RDCs of the alignments as illustrated in Figure 8.1.

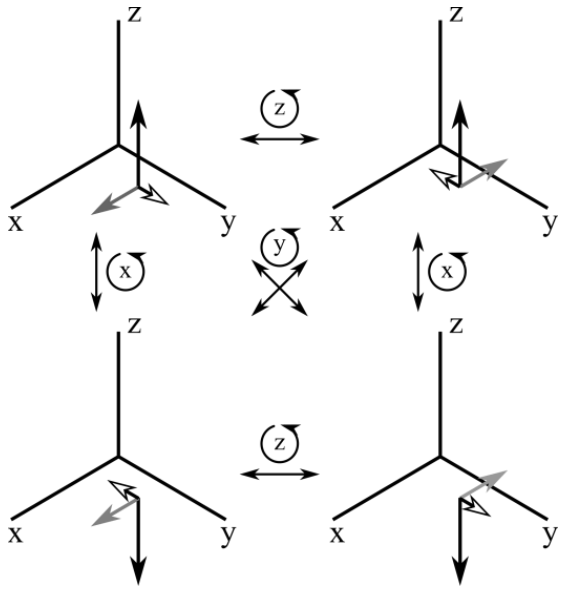



**Figure 8.1. The four positions of a molecule (here represented by three vectors) which will give equal RDC values from tensor RDC calculations. All of the indicated rotations are 180° around the axis.**

In Figure 8.2 left, the direction of the magnetic field is along the z-axis, and the length of the *a* vector will serve as the degree of alignment ($A_a$). But the direction of the magnetic field will also be along the x-axis to some degree, and the length of *b* determines how much it is aligned here ($A_r$). The probability, or degree, of the shift is given by the rhombicity ($R$) given in equation (8.1).

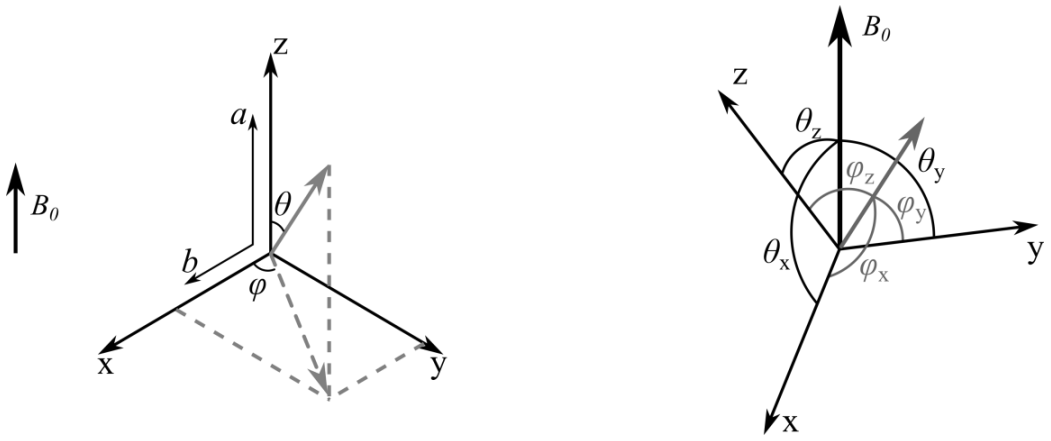



**Figure 8.2. Left: The scaling vectors and angles involved in the extended formula for tensor RDC calculations, equation (8.1). Right: The molecular axis compared to the direction of the magnetic field, including the angles used in SVD fitting. The angles $\theta$ are the angles between the molecular axis and the magnetic field, and the angles $\varphi$ are the angles between an internuclear vector and the molecular axis, see equations (8.1) to (8.12).**

Without introducing the rhombicity, the theoretical fit between experimental and back-calculated data would be worse for most alignments since some averaging is present in the experimental data, and the compound would not be specifically placed in space but only by an angle to the magnetic field with free rotation around the magnetic field vector. An experimental dataset which back-calculates to a structure with a low rhombicity is thus averaged less along the x-axis than a structure with a high rhombicity.

$$A_a = \frac{S_{zz}}{2} \tag{8.2}$$

$$A_r = \frac{1}{3}\left(S_{xx} - S_{yy}\right) \tag{8.3}$$

$$R = \frac{A_r}{A_a} = \frac{2}{3}\frac{S_{xx} - S_{yy}}{S_{zz}} \tag{8.4}$$

To establish a tensor the equations (8.1)-(8.4) are used to produce an order matrix as a representation of the anisotropic averaging leading to dipolar couplings, given in equation (8.5) and (8.6) Figure 8.2, right is a good reference to keep track of the angles used in the following.

$$A = \begin{bmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{zz} \end{bmatrix} \tag{8.5}$$

$$S_{ij} = \langle 3\cos\theta_i \cos\theta_j - k_{ij} \rangle \tag{8.6}$$

Where $k_{ij}$ is the Kronecker delta ($k_{ij} = 1$ if $i = j$ and $k_{ij} = 0$ if $i \neq j$). The matrix is symmetric ($S_{ij} = S_{ji}$) and traceless ($S_{xx} + S_{yy} + S_{zz} = 0$). Thus only five elements are needed to obtain the full matrix *A*, chosen to be $S_{yy}$, $S_{zz}$, $S_{xy}$, $S_{xz}$ and $S_{yz}$. The five elements translate to three angles (which need to be

non-parallel) the alignment strength and the rhombicity. The equations (8.5) and (8.6), if used directly on a 3D structure, suppose a situation where the alignment frame and the molecular frame have an identical orientation which is not generally true. To correlate the molecular frame to the alignment frame, the equation is thus extended into equation (8.7) and further to (8.8) where $\varphi_i$ is angle between the two frames.

$$D^m = D^m_{max} \sum_{ij=x,y,z} S_{ij} \cos\varphi_i^m \cos\varphi_j^m \tag{8.7}$$

$$D^m_{red} = \frac{D^m}{D^m_{max}} = \sum_{ij=x,y,z} S_{ij} \cos\varphi_i^m \cos\varphi_j^m \tag{8.8}$$

Where $D^m{}_{max}$ is $-\mu_0\gamma_I\gamma_S\hbar r^{-3}/(16\pi^2)$. This equation is then used to setup a set of linear equations, as seen in (8.9). Note that the $\varphi$ angles are always known from a 3D structure and the matrix is easily established.

$$\begin{bmatrix} D^1_{red} \\ \vdots \\ D^n_{red} \end{bmatrix} = \begin{bmatrix} S_{yy} \\ S_{zz} \\ S_{xy} \\ S_{xz} \\ S_{yz} \end{bmatrix} \begin{bmatrix} cos^2\varphi^1_y - cos^2\varphi^1_x & cos^2\varphi^1_z - cos^2\varphi^1_x & 2cos\varphi^1_x cos\varphi^1_y & 2cos\varphi^1_x cos\varphi^1_z & 2cos\varphi^1_y cos\varphi^1_z \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ cos^2\varphi^n_y - cos^2\varphi^n_x & cos^2\varphi^n_z - cos^2\varphi^n_x & 2cos\varphi^1_x cos\varphi^n_y & 2cos\varphi^1_x cos\varphi^n_z & 2cos\varphi^1_y cos\varphi^n_z \end{bmatrix} \tag{8.9}$$

For $n > 5$ the system will be overdetermined, while it will be unsolvable for $n < 5$. As a consequence, RDCs of five non-parallel vectors are needed to describe a system using this method. If $n = 5$ the system is solvable as five equations with five unknowns, but this situation is not particularly interesting as one could in principle fit almost anything to "perfection". Thus SVD was introduced as a mean to solve the overdetermined set of linear equations, and to find the order tensor with the best overall fit to the data in a least square sense.

It is used that the $N\times M$ matrix $[B]$ in equation (8.10), from the matrix in (8.9), may be described by the product of an $M\times N$ matrix $[U]$, an $N\times N$ diagonal matrix $[W]$, with non-negative diagonal elements, and the transpose of a $N\times N$ orthogonal matrix $[V]$, which allows the definition of $[B]^{-1}$ as can be seen in equation (8.11).

$$\begin{bmatrix} D^1_{red} \\ \vdots \\ D^n_{red} \end{bmatrix} = \begin{bmatrix} S_{yy} \\ S_{zz} \\ S_{xy} \\ S_{xz} \\ S_{yz} \end{bmatrix} [B] \tag{8.10}$$

$$[B] = [U] \begin{bmatrix} \omega_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \omega_n \end{bmatrix} [V^T] \;\leftrightarrow\; [B]^{-1} = [V][diag(1/\omega_i)][U]^T \tag{8.11}$$

This may then be used to solve equation (8.9) as equation (8.12).

$$\begin{bmatrix} S_{yy} \\ S_{zz} \\ S_{xy} \\ S_{xz} \\ S_{yz} \end{bmatrix} = \begin{bmatrix} D^1_{red} \\ \vdots \\ D^n_{red} \end{bmatrix} [V][diag(1/\omega_i)][U]^T \tag{8.12}$$

SVD is implemented as a function in mathematical programs such as Matlab®, making the decomposition easy.[247]

The result is a least square fit of the experimental data to the 3D structure, which will rotate the structure into the alignment frame and scale the RDC values according to the experimental data.[202] It is thus a back-calculation method where a 3D structure is needed to determine the order tensors. For rigid structures SVD works very well and is a powerful, albeit simple, tool in macromolecular and small molecule work alike.[187,195,197,248]

The drawback of this method is evident only if investigating flexible compounds. The arising problems are caused by establishing order tensors in the first place and are not a flaw in the SVD fitting as such. Since order tensors will determine a rotation of the structure the implementation for multiple structures is not straightforward but may still be approached. In general two methods have been applied; the multi conformer multi tensor (MCMT) and multi conformer single tensor (MCST) methods.[248] The MCMT method builds on the assumption that one may describe the orientation of flexible systems by obtaining order tensors for each conformer. As this method involves multiple order tensors, 5*n* RDCs are needed *per conformation*, with *n* being the number of conformers of the system.[202] This approach is thus rarely feasible for flexible small molecules.[202,248]

### 8.1.1 Multi conformer single tensor

The MCST method assumes that a single tensor is sufficient to describe the alignment of all conformers.[248] In other words it is assumed, that all conformers are aligned equally in space – or at least parts of the conformers. This may be a good approximation in certain situations, when the structural flexibility is limited to small groups in regard to an overall structure or to small movements. This is an obvious limitation, since a reference frame needs to be established and all structures are linked to this reference frame when rotated into the alignment frame.[248] The resulting back-calculated RDCs are thus dependent on the original overlay of the input structures, and parts of the molecule need to be identified that are independent in the orientation in space between the conformers. While a rigid part may be identified for the majority of organic compounds (if not a different set of problems will be more pressing in 3D structure determination) it may be harder to identify a part of a molecule that, independent of conformation, is situated evenly in space compared to the rest of the structure. A general approach to identify the optimal overlay of structures has yet to be reported. An approach could be to overlay vectors with the largest absolute RDCs, as these are probably the most static in their orientation to $B_0$. Whether this is better than e.g. identifying large static surfaces of the compound is not known, and was not thoroughly investigated.

The methodology is thus limited to specific cases and over-interpretation is deceptively easy, which will be showed *vide infra*. It may be implemented in different ways; here the focus will on be the implementation utilized in the program MSPIN.[204] Here the population is iteratively fitted to the experimental RDCs, by constructing alignment tensors from the populations and the evaluating the correlation of experimental and calculated RDCs.

### 8.1.2 RDCs of methylene and methyl groups

When dealing with overlying resonances such as methylene groups that do not split and methyl groups, various methods have been proposed to back-calculate the RDCs.

$^1D_{CH3}$ values are usually obtained from spectra and, while not immediately useful, are converted to the associated $^1D_{CCH3}$ coupling constant, which is easier to implement in the given types of back-calculations. This is achieved by assuming a 3-jump model and that the methyl group is a perfect tetrahedral in equation (8.13).[249,250] This assumption makes a conversion from a theoretical CH vector to the actual CC vector possible using equation (8.14).

$$^1D_{CH_3} = \frac{3\cos^2\varphi - 1}{2} D_{\parallel} = -\frac{1}{3} D_{\parallel} \quad \textbf{(8.13)}$$

$$^1D_{CCH_3} = \frac{\gamma_C}{\gamma_H} \frac{{r_{CH}}^3}{{r_{CC}}^3} D_{\parallel} \rightarrow D_{\parallel} = -\frac{3\gamma_C}{\gamma_H} \frac{{r_{CH}}^3}{{r_{CC}}^3} \, ^1D_{CH_3} \quad \textbf{(8.14)}$$

Where $\varphi$ is the angle of the CH vector and the rotation axis of the methyl group, $D_{\parallel}$ is a virtual CH vector pointing in the direction of the CC vector, $\gamma$ is the gyromagnetic ratio and $r$ is the length of the corresponding bonds.

For methylene groups it has been shown that if an average coupling constant is extracted, the data may still be used in the fitting of experimental data to 3D structures, by calculating the theoretical average from the individual CH vectors.[208]

## 8.2 Tensor free calculations of RDCs

Tensor free calculation of RDCs, also dubbed the ϑ-method, has been shown to work well for macromolecules but has not directly been adopted in the small molecule community.[251,252] In a publication by Luy *et al.* in 2015, a resembling although different methodology is utilized.[245] For macromolecules the ϑ-method is used implicitly in the 3D structure generation as constraints.[251,252] This is not necessarily wanted for small molecules due to similar reasons as discussed in e.g. Section 3.3.1 for NOE calculations: In short, the use of constrains may limit or disturb the conformational space.

We thus set out to translate the method into one more suitable for conformational investigation of small molecules. In the following description this method is called the $\theta$-method, for distinction and recognition of inspiration.[251,252]

## 8.3 The $\theta$-method

The $\theta$-method is simplistic in its origin; it utilizes only the standard equation for dipolar couplings, equation (8.15) reiterated from equation (5.2) Section 5.1, with no tensors to correlate the alignment frame to the molecular frame, as shown in Figure 8.2.

$$D_{IS} = -\frac{\hbar \gamma_I \gamma_S \mu_0}{16\pi^2} \langle \frac{1}{r_{IS}^3} (3\cos^2\theta_{IS} - 1) \rangle \quad \textbf{(8.15)}$$

This equation may be shortened to (8.16).

$$D_{IS} = D_{max}^{IS} D_a \langle 3\cos^2\theta_{IS}(R) - 1\rangle = D_{max}^{IS} D_a \int P(R)(3\cos^2\theta_{IS}(R) - 1)dR \quad \textbf{(8.16)}$$

$R$ covers external (rotational) and intramolecular (conformational) motion, $D_a$ is the degree of alignment and $\kappa$ is the constants of (8.16) assuming that only $^1D_{CH}$ data is used and $r$ is constant. If this is not the case the equation is expanded to (8.17).

$$D_{IS} = D_{r,max}^{IS} D_a \left\langle \frac{3\cos^2\theta_{IS}(R) - 1}{r(R)^3} \right\rangle = D_{r,max}^{IS} D_a \int P(R) \left( \frac{3\cos^2\theta_{IS}(R) - 1}{r(R)^3} \right) dR \quad \textbf{(8.17)}$$

This may in principle be solved if all conformers and a rotational preferences are known. Luy *et al.* showed this approach worked on a system where strychnine was simulated in the presence of a polymer strand which induced overall alignment.[245]

Instead of including explicit orienting media in the simulations or orienting the molecular frame to an alignment frame, a novel approach was taken; all conformers from a simulation are rotated in space, using rotation matrices as in (8.18).

$$Rot_z(\varphi) = \begin{bmatrix} cos\varphi & -sin\varphi & 0 \\ sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix}, Rot_x(\omega) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\varphi & -sin\varphi \\ 0 & sin\varphi & \cos\varphi \end{bmatrix} \quad \textbf{(8.18)}$$

To get full rotational sampling in 3D space compared to one axis (here z), a combination of only two matrices are needed; x and y, x and z or y and z (for similar reasons as outlined in Figure 8.1). Here, x and z were chosen, so that the first rotation is around the z axis, which is also the chosen $B_0$ axis, and the second rotation is around the x axis, as illustrated in Figure 8.3. In practice, one of the rotations only needs to be 180° as to avoid repeating magnetic field angles, while the other is 360°.

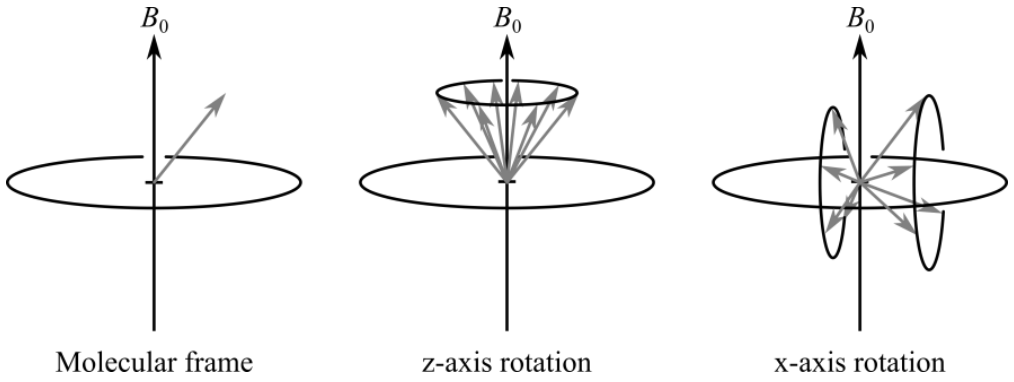



**Figure 8.3. An illustration showing the effect of z- and subsequent x-rotation on an internuclear vector. While the angle to the magnetic field is equal for all vectors after z-rotation, the angle is different for the vectors after equal x-rotation (unless a vector is parallel to z-axis).**

The angle of the RDC vectors from the rotated coordinates to the magnetic field was then determined along with the length of the vector, which in most cases is constant since $^1D_{CH}$-coupling constants are used. From the length and angle RDCs of all conformers and rotations are easily calculated after $D_a$ is determined, *vide infra*. The RDC data sets of the rotated structures are then fitted to the data iteratively, by minimizing the *Q*-factor (*Q*) in (8.19), using only the conformations and rotations that lead to an increased correlation of the found average RDCs to the experimental.

$$Q = \frac{\sum_i D_i^{calc} - D_i^{exp}}{\sum_i D_i^{exp}}, D_i^{calc} = \frac{\sum_n D_{i,n}^{calc}}{n} \quad \textbf{(8.19)}$$

Thus, all conformers may be aligned differently, if the data permits. This is at a first glance a possibly problematic approach with multiple open questions:

1. Would the approach give reasonable fits between experimental data and 3D structures?
2. Would the calculation procedure make it possible to discern stereoisomers?
3. And most importantly, will this method fit populations of flexible compounds?

The easiest way to verify the method is to test it on a compound where the answer is already known and return to rigid compounds to establish prove of concept. Multiple tests were thus conducted on multiple data sets of strychnine, *vide infra*.

If long-range RDCs were needed the length of the vector of the different conformations need to be calculated and averaged as well. This was in practice easily implemented by incorporating the distance information into the theoretical RDCs, as in equation (8.17).

## 8.4 Estimating $D_a$

As mentioned above the strength of alignment needed to be determined or estimated. Here inspiration from macromolecular literature was used.[253] The difference of the approach taken here lies in the fact that $D_a$ is not used for tensor determination and is not split into an axial ($D_a$) and a rhombic component ($D_r$) and the rhombicity is thus not estimated as it is not needed due to implicit inclusion in tensor free calculations. Even though the two $D_a$ parameters are used differently, their contribution to the calculations is equal in giving the maximum possible values of the RDCs, and they may be cautiously compared.

Initially it is assumed that the vectors in the molecules are non-parallel and of equal length (neither assumptions are strictly necessary, but simplify the following)[VII], and that at least one random vector is either parallel or perpendicular to the magnetic field, $B_0$. That particular vector would then have the minimum ($D_{exp,min}$) or maximum ($D_{exp,max}$) experimental RDC respectively from equation (8.15) or (8.16). From the equation it is evident that the equations in (8.20) are true, as the constant $k$ covers all other contributions to the RDCs than the angle to the magnetic field.

$$\begin{aligned} D_{exp,min} &= -k\left(3\cos^2\theta_{IS} - 1\right) = -k\left(3\cos^2 0 - 1\right) = -2k \\ D_{exp,max} &= -k(3\cos^2\theta_{IS} - 1) = -k(3\cos^2\pi - 1) = 1k \\ D_{exp,max} &= -D_{exp,min}/2 \\ R_D &= -D_{exp,min}/2D_{exp,max} \end{aligned} \quad \textbf{(8.20)}$$

[VII] If all vectors are parallel, $D_a$ may not be estimated, but the angles between internuclear vectors (which are already known) may in principle be solved and the molecule will be fitted with all internuclear vectors either parallel of perpendicular to the magnetic field, dependent on the sign of the RDCs.

From these equations $D_a$ may be estimated; if $R_D > 1$ then $D_a$ is $-D_{exp,min}/2$, and if $R_D \leq 1$, $D_a$ is $D_{exp,max}$, divided by $D_{max}$ of the given vector. The determined $D_a$ is thus the minimum $D_a$ that could possibly lead to the experimental data. It is crude, but may be refined later if needed by up- or down-scaling $D_a$. It is assumed that $D_a$ is equal for all conformers. This approximation is actually a prerequisite for most RDC back-calculation theory; it translates to assuming that the interaction with the alignment media is independent of conformation, and thus does not influence the conformational average. Since that requirement is necessary to obtain any meaningful conformational data from RDCs this assumption is hopefully upheld. There is a larger possibility of finding a vector that is perpendicular to the magnetic field, easily realized by the fact that the probability of finding a vector with an angle $\theta$ to the magnetic field is proportional to $\sin\theta$.[11,253] This is generally not considered in the following, since the value of $D_a$ is scaled, but it could be introduced to the calculation.

## 8.5 Implementation

The $\theta$-method was implemented in Matlab®. The overall methodology of the script is described above or found in Appendix A7. The script handles RDCs, NOEs and *J*-coupling constants, individually or in combination, given input files and a structure file with one or more structures (.pdb, .mol2 or .sdf supported). Examples of the different inputs are given in the experimental section.

## 8.6 Results

In this section the $\theta$-method is utilized to back-calculate RDCs of several compounds and the results are compared to SVD calculations using the MCST approximation when appropriate. But first a couple of questions needed to be answered in order to establish a proof of concept of the method.

### 8.6.1 Proof of Concept - Strychnine

Several questions needed to be answered, in order to evaluate the applicability of the $\theta$-method, and are addressed below. Since the method is based on existing methods, the proof of concept part is more a study in the limitations than whether the method is possible, though that is also investigated. To achieve proof of concept the structure of strychnine will be utilized below. Strychnine is rigid with multiple $^1D_{CH}$ vectors, and is thus generally utilized for many purposes. Two datasets of strychnine were investigated with correlating results. One of the sets is used as an example in the following.

### Initial fitting of strychnine

The structure of strychnine was easily fitted to the RDC data using either SVD or the $\theta$-method, as seen in Table 8.1.

**Table 8.1. Experimental versus calculated RDCs for the major conformer of strychnine. Calculated by SVD in MSPIN and the $\theta$-method. Aligned in PDMAA (Polymer 8.1).**

| Nuc 1 | Nuc 2 | Exp. | $\theta$-method | SVD |
|---|---|---|---|---|
| C1 | H1 | 26.6 | 26.3 | 26.2 |
| C2 | H2 | 13.8 | 13.7 | 13.9 |
| C3 | H3 | 23.8 | 20.1 | 20.3 |
| C4 | H4 | 28.3 | 26.6 | 26.6 |
| C8 | H8 | -14.4 | -14.1 | -14.4 |
| C11 | H11a | 12.3 | 13.8 | 13.7 |
| C11 | H11b | -25.8 | -28.6 | -28.9 |
| C12 | H12 | -28.4 | -29.3 | -29.3 |
| C13 | H13 | -4.3 | -4.5 | -4.7 |
| C14 | H14 | -27.9 | -29.1 | -29.0 |
| C15 | H15a | 7.0 | 5.6 | 5.5 |
| C15 | H15b | 7.5 | 7.4 | 7.3 |
| C16 | H16 | 15.2 | 14.2 | 14.2 |
| C18 | H18a | 1.6 | -0.4 | -0.5 |
| C18 | H18b | 14.7 | 13.6 | 13.6 |
| C20 | H20a | -16.0 | -14.5 | -14.6 |
| C20 | H20b | -4.1 | -3.4 | -3.6 |
| C22 | H22 | 0.7 | 3.7 | 3.6 |
| | *Q*-factor | | 0.092 | 0.092 |

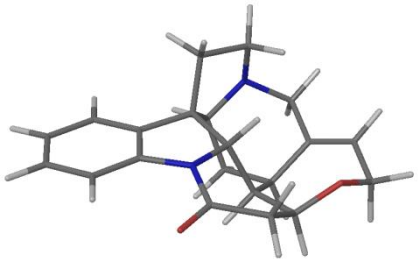


**Figure 8.4. The major conformer of strychnine**

The resulting back-calculated RDCs were very much alike with an RMSD of 0.13 Hz and a MAE of 0.06 Hz between the two methods. While this was reassuring the work relied heavily on the following. The general alignment is compared in Figure 8.5, but as the averaging is implicit in SVD and explicit in the $\theta$-method, the comparison is of only the most abundant alignment.

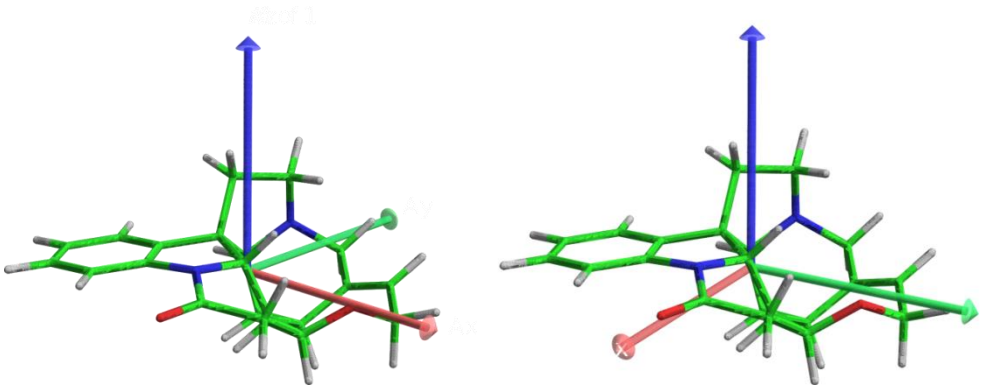


**Figure 8.5. The alignment frame from and SVD fitting of strychnine (left) and the molecular frame of the most abundant alignment to the magnetic field from a $\theta$-method fitting. Both alignments have the z-axis aligned with the magnetic field. Note that the x- and y-axis are without influence in the $\theta$-method, and the alignments are almost identical. Aligned in PDMAA (Polymer 8.1).**

### Rotation angle $\varphi$

The first potential drawback of the method is caused by the fact that a rotation must be defined, and thus the back-calculation of RDCs is based on a finite number of rotations. It is not possible to

make the rotation angle infinitely small due to computer memory and time of calculations. In comparison the SVD method rotates the structure into the linear best fit with a theoretical infinite rotational resolution. Conditions were tested to establish that it was possible to find a rotation angle which give reliable results and is computationally viable.

To investigate this, the 3D structure of the major conformer of strychnine was used. When optimized this structure has a random orientation in space and randomly 100 structures were generated from 100 random rotation angles of the structure along the z- and x-axis (using a rotational resolution below 0.0001 degrees). The 100 randomly chosen structures were rotated by 90, 60, 30, 10, 1, 0.5 or 0.1 degrees and back-calculated RDCs compared to the experimental data. The summary of this may be found in Figure 8.6 and Table 8.2.

The lowest rotation angle used in the further investigation was 0.1 degree which is a practical decision; to decrease an order of magnitude from 0.1 degree one would need ~227 GB of RAM, way above standard computer setups (and even clusters).[VIII] This could probably be alleviated by saving and loading the structures continuously in the Matlab® script, but this was not a feasible solution due to increased calculation times.

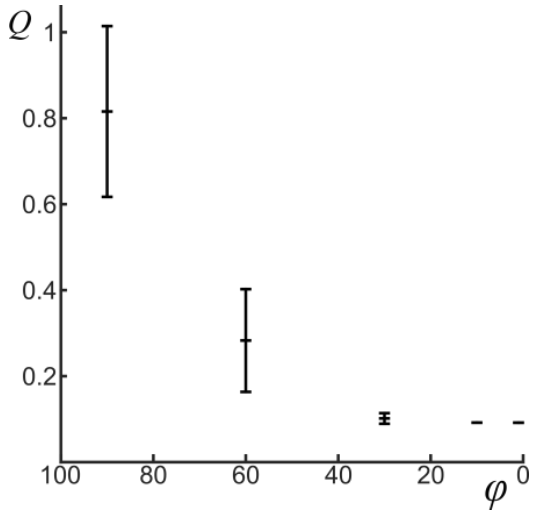



**Figure 8.6. *Q*-factor of an RDC fitting of strychnine using different rotation steps (*φ*) in degrees. Mean and standard errors (error bars) shown. Data set used for strychnine is Polymer 8.1.**

**Table 8.2. Statistics of the *Q*-factor of an RDC fitting of strychnine using different rotation steps (*φ*) starting from 100 different rotations. Data set used for strychnine is Polymer 8.1.**

| *φ*/° | Mean | St. dev. | Median | Min | Max |
|---|---|---|---|---|---|
| - | 1.6290 | 0.4154 | 1.6950 | 0.4767 | 2.2857 |
| 90 | 0.8156 | 0.1986 | 0.8910 | 0.2060 | 0.9981 |
| 60 | 0.2829 | 0.1195 | 0.2835 | 0.0934 | 0.5357 |
| 30 | 0.1016 | 0.0123 | 0.0963 | 0.0919 | 0.1468 |
| 10 | 0.0919 | 0.0002 | 0.0919 | 0.0916 | 0.0924 |
| 1 | 0.0918 | 0.0002 | 0.0917 | 0.0915 | 0.0921 |
| 0.5 | 0.0917 | 0.0002 | 0.0916 | 0.0915 | 0.0921 |
| 0.1 | 0.0917 | 0.0001 | 0.0916 | 0.0915 | 0.0921 |

It should be addressed that a rotation of 10 or even 30 degrees fitted the data just as well as lower rotational steps. It is unlikely that a rotation of 30 degrees will produce a structure that is situated rightly in the alignment frame, and the generally low *Q-factor* is attributed to the fact that the rotational average is optimized. It is thus possible to construct data from multiple aligned structures which averages in line with the observed data. If a rotation by 1 degree or lower is utilized, the possibility to be in the alignment frame is vastly increased, and the minor averaging needed may

[VIII] This is a result of the script where all rotations of all vectors are saved in matrices. Estimated as RAM(0.1° rot.)×10×10.

even correlate well to what is actually observed in experiments. Due to the vast increase in computational requirements when decreasing the rotation step size by an order of magnitude, a rotation of 1 degree was used to perform the back-calculations and fittings in the remainder of this thesis.

**$D_a$ estimations**

To do the calculations as outlined in (8.16) and (8.17) it was a prerequisite that $D_a$ may be estimated. The theoretical approach was already explained in (8.20). In order to test whether the approach was viable, $D_a$ was scaled by a constant $S_{Da}$ in the fitting to experimental data and the resulting $Q$-factor was evaluated. An example of this, using strychnine again, is seen in Figure 8.7.

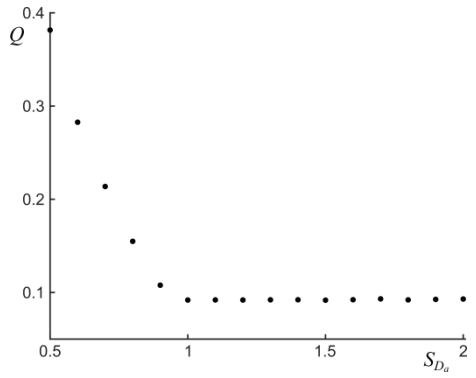



**Figure 8.7. $Q$-factor of an RDC fitting of strychnine using a $D_a$ obtained by the approach outlined in Section 8.4 and scaled by $S_{Da}$. Data set used for strychnine is Polymer 8.1.**

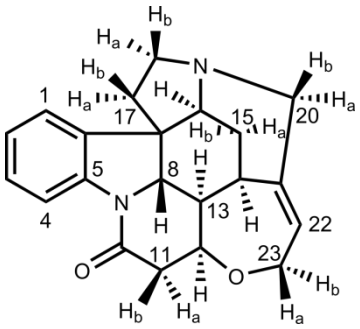
The assumptions used in estimating $D_a$ seemed to be valid for strychnine in the experimentally obtained alignment frame, as an increase in $D_a$ did not lead to a better correlation to the experimental data. A decrease in $D_a$ led to a worse correlation, in good correlation to the fact that the larger observed RDCs, independent of sign, are no longer obtainable. The fact that $S_{Da} = 1$ was the optimum means that a vector was situated perpendicular or parallel to $B_0$. In the current case it was the vector of C4-H4 which was determined to be perpendicular to the magnetic field. This will not be true for all structures or alignments, and thus each aligned dataset was treated as above to determine if the approximated $D_a$ value is appropriate for the data, and the value was scaled accordingly for the specific dataset if needed.

**Differentiation of stereoisomers**

Strychnine has previously been used for RDC calculations and multiple studies have shown that it is possible to assign the diastereotopic protons of strychnine by RDCs.[187,188,208] To probe whether the $\theta$-method may be used to differentiate stereoisomers, the diastereotopic protons of strychnine were switched, fitted to the data set and compared to the right assignment. This was compared to the results of a fitting of the same data using SVD as implemented in the commercial program MSPIN. The comparison is found in Table 8.3.

**Table 8.3. Comparison of the distinction of right and wrong assignment of the diastereotopic protons in strychnine by the *θ*-method and SVD. The diastereotopic protons are switched as stated, and may be compared to the right assignment in the top row. Aligned in PDMAA (Polymer 8.1).**

| | ***Q-factor*** | | **Ratio** | |
|---|---|---|---|---|
| Switch # | *θ*-method | SVD | *θ*-method | SVD |
| Right | 0.093 | 0.087 | 1.0 | 1.0 |
| Dia-11 | 0.688 | 0.565 | 7.4 | 6.5 |
| Dia-15 | 0.097 | 0.088 | 1.0 | 1.0 |
| Dia-18 | 0.243 | 0.209 | 2.6 | 2.4 |
| Dia-20 | 0.191 | 0.180 | 2.1 | 2.1 |
| Dia-23 | 0.523 | 0.458 | 5.6 | 5.3 |




The distinction was just as good for the *θ*-method as for SVD. The reason that the diastereotopic protons of C-15 were indistinguishable is due to very similar observed RDCs, and thus lies in the orientation of the molecules in the alignment media and not the method. Since the *θ*-method may easily be used in tandem with *J*-coupling constant and/or NOE data, the combined data was compared to 3D structures by addition of RMSDs between all experimental and theoretical data, allowing the differentiation of H-15a/H-15b, see Table 8.4. This may of course also be done for SVD fittings by separate investigation of the other parameters, but this approach slightly more cumbersome.

**Table 8.4. RMSD of RDC data, alone or coupled to NOE and/or *J*-coupling constant data for the diastereotopic protons at C15 of strychnine if assigned right or wrong. Only NOEs and *J*-coupling constants involving H15a/b were used.**

| | **RMSD** | |
|---|---|---|
| | Right | Dia-15 |
| RDC | 1.247 | 1.266 |
| RDC/*J* | 1.853 | 3.207 |
| RDC/NOE | 1.287 | 2.247 |
| RDC/J/NOE | 1.877 | 4.186 |

## Determining low level conformers of strychnine

A low level conformer of strychnine was previously published from NOE data, and recently attempts were made to use the MCST approach to determine the population of the two conformers from RDCs.[30,248] The attempt showed varying results and as a consequence the performance of the *θ*-method was tested. Data from the study above as well as another dataset used in Section 6 were tried which may be compared to the literature.[248]

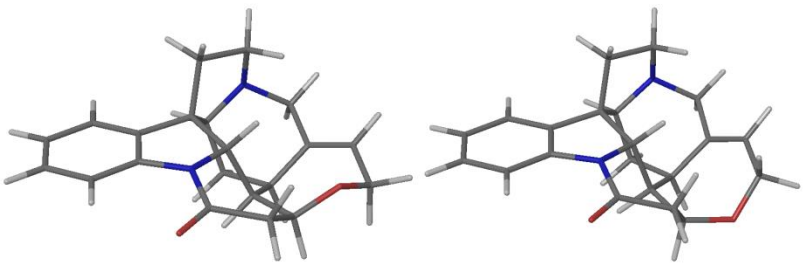


**Figure 8.8. The major (left) and minor (right) conformer of strychnine. It is noted that a third conformer has been presented, but this conformer is present well below one percent and not used for comparison.[30,248,254]**

In line with the MCST SVD fittings in the literature, the studies here also concluded that it was hard to obtain a meaningful conformational population for the structures of strychnine by the first set of RDCs. The two methods gave more or less exactly the same populations from the first dataset, using MCST and the *θ*-method alike, with an overestimation of the minor conformer as given in Table 8.5. For the second dataset the population is the same when using the *θ*-method while the MCST performs significantly better. Compared to the literature study by Thiele *et al.* the populations were in good agreement as the population of conformer 1 was populated from around 0.8 to 1 for multiple literature datasets.[248]

**Table 8.5. Populations of the two conformers of strychnine, using only one bond RDCs or including long-range RDCs. Scaled SVD covers the usage of "SVD Hz scaling" option in MSPIN, where the larger RDCs contribute more to the fit. This was needed to obtain good *Q*-factors when using the full datasets. (Dataset 1: Polymer 8.1, dataset 2: Polymer 6.2).**

| | $^1D_{CH}$ | | | $^1D_{CH}$+$^nD_{HH}$ | | |
|---|---|---|---|---|---|---|
| | Conf. 1 | Conf. 2 | *Q*-factor | Conf. 1 | Conf. 2 | *Q*-factor |
| | | | Dataset 1 | | | |
| SVD | 0.79 | 0.21 | 0.081 | 1.00 | 0.00 | 0.223 |
| SVD, scaling | 0.81 | 0.19 | 0.080 | 0.96 | 0.04 | 0.168 |
| *θ*-method | 0.81 | 0.19 | 0.072 | 0.88 | 0.12 | 0.164 |
| | | | | | | Dataset 2 |
| SVD | 0.96 | 0.04 | 0.079 | 0.98 | 0.02 | 0.296 |
| SVD, scaling | 0.96 | 0.04 | 0.079 | 1.00 | 0.00 | 0.119 |
| *θ*-method | 0.84 | 0.16 | 0.073 | 0.95 | 0.05 | 0.118 |
| NOE[30] | 0.98 | 0.02 | | 0.98 | 0.02 | |

To test whether the population fit would be improved by utilizing the additional vectors from the $S^3$ HMBC *homo*, see Chapter 6, back-calculation were performed which included these data. The reasoning behind a possible better fit was that the additional data involve internuclear vectors involving the protons 23a/23b, which will differ in length between the conformations. Due to the $r^{-3}$ dependence of the resulting RDC this was thought to greatly influence the resulting populations from the data. Indeed, the inclusion of long-range RDC data generally improved the populations for both methods, as seen in Table 8.5, compared to the expected population.

The MCST approach seemed to be slightly better for this simple system. This could be attributed to the fitting procedure, where MCST uses iterative population fitting, while the *θ*-method is forced to rely on an additive best fit approach due to the amount of "structures" generated from the rotations. It is also beneficial for MCST that the changes in the overall structure were minor between the two conformations, and thus the overlay of the structures was easily determined. In conclusion both methods seemed to be inferior to an NOE approach, which was not surprising due to the $r^{-6}$ dependence leading to a huge impact by small changes in distances on the NOE intensities, while also minimizing contributions from errors.

To test the scope of the *θ*-method other organic compounds were tested by fitting of experimental RDCs to multiple conformers. These will be presented below.

### 8.6.2 Reserpine

Reserpine is a plant alkaloid which features a linear pentacyclic system, see figure in Table 8.6. At least one strained ring conformation is present at all time which was thought to lead to flexibility by differences in ring conformation between conformers. Unlike strychnine, where the molecular structure is very rigid and very few structures were identified, reserpine had have many different conformations in the simulation, though a single conformer of the ring system was identified which is populated approximately 90 % from force field energies. The evaluation of many conformers could be problematic in an SVD fitting in MSPIN as the computational time increases rapidly with the number of structures. One approach could be to determine the relative energies from more accurate DFT calculations, or trust those of the force field method and remove conformers below a certain threshold, and thereby limit the amount of structures. In Figure 8.9 the conformational space of reserpine is illustrated, with focus on the pentacyclic system.

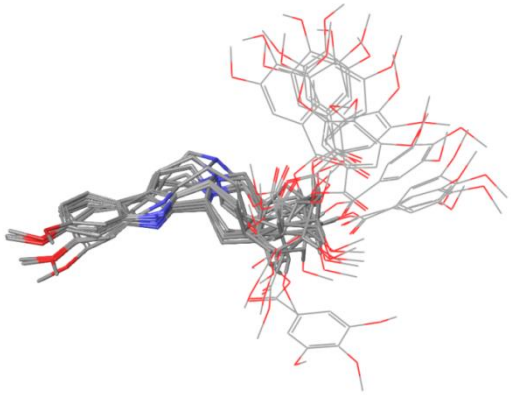
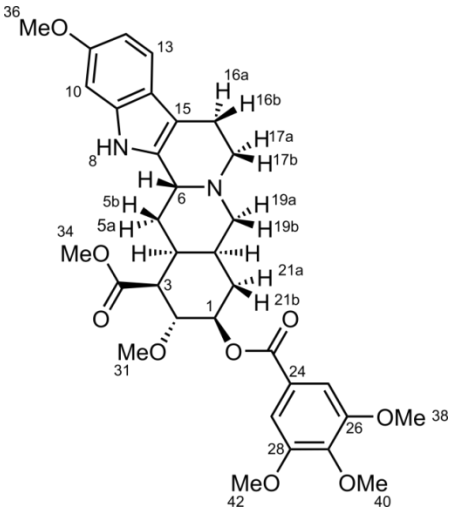


**Figure 8.9. Conformational space covered by the pentacyclic ring of reserpine.**

All conformers were included in the fitting procedure disregarding any energy differences determined in the FF minimization. This should test the prowess of the $\theta$-method and see if the correct conformer was obtained. The correlation obtained from the $\theta$-method, and those using SVD, between experimental and back-calculated data, is found in Table 8.6. For the first SVD a single structure, representing the pentacyclic system with the lowest energy was used. For the second structures that represented the conformational averaging of the pentacyclic system were used, due to the amount of structures else needing fitting. The $\theta$-method utilized structures covering a much larger conformational space (114 conformers). Both methods identified the correct major conformer, but some observations indicated that the SVD method was problematic: The correlation to experimental data was markedly worse and the $D_a$ for the multiple-conformers fitting was much higher than in the two other cases. This may be contributed to the overlay of the structures, where the heavy atoms in the pentacyclic system were used. It was not possible to establish a better overlay, so this may not be the problem. Also the limited number of input structures could have been problematic. Still, the ability to input all conformations in the $\theta$-method, rendered all post-simulation pre-fitting considerations, such as overlay of atoms, unnecessary, as a big dataset with many structures could easily and quickly be fitted to experimental data without prior 3D structural assumptions.

**Table 8.6. Experimental and back-calculated RDCs for reserpine in Hz. Calculated by SVD in MSPIN and the *θ*-method. The data from SVD was compared to a single structure or conformers representing the core pentacyclic system and 3-jump models of the rotatable groups. $D_a$ is scaled by 1.3 in the *θ*-method. (Polymer 8.2). Right: Structure of reserpine.**

| C | H | Exp | *θ*-method | $SVD_{single}$ | $SVD_{mult}$ |
|---|---|---|---|---|---|
| 1 | 1 | -36.3 | -36.1 | -33.4 | -35.2 |
| 2 | 2 | -35.5 | -36.3 | -35.7 | -37.1 |
| 3 | 3 | -38.5 | -37.5 | -30.7 | -24.3 |
| 4 | 4 | 12.5 | 14.5 | 13.9 | 16.1 |
| 5 | 5a | -18.9 | -20.4 | -22.9 | -18.2 |
| 5 | 5b | 22.5 | 20.2 | 13.7 | 20.7 |
| 6 | 6 | -28.5 | -28.9 | -27.3 | -29.2 |
| 10 | 10 | 2.3 | 2.1 | 2.9 | 2.8 |
| 12 | 12 | 27.4 | 29.7 | 30.6 | 40.4 |
| 13 | 13 | 2.4 | 1.7 | 2.5 | 2.7 |
| 16 | 16a | -55.7 | -57.0 | -41.3 | -56.0 |
| 16 | 16b | 36.9 | 33.6 | 31.5 | 23.7 |
| 17 | 17a/17b | -37.6 | -31.0 | - | - |
| 19 | 19a | 15.8 | 15.4 | 12.5 | 6.5 |
| 19 | 19b | -17.6 | -17.5 | -22.0 | -20.2 |
| 20 | 20 | -33.4 | -33.2 | -30.8 | -35.7 |
| 21 | 21a | -27.5 | -27.1 | -26.5 | -23.8 |
| 21 | 21b | -22.9 | -23.4 | -24.0 | -16.0 |
| 25/29 | 25/29 | -16.8 | -15.7 | - | - |
| 31 | 31 | 0.4 | 0.5 | 8.9 | -7.4 |
| 34 | 34 | 7.4 | 7.6 | 6.0 | 0.9 |
| 36 | 36 | -1.5 | -1.8 | -0.1 | -12.2 |
| 40 | 40 | -2.0 | -3.4 | -1.9 | -3.9 |
| | | *Q*-factor | 0.055 | 0.165 | 0.259 |
| | | *Q* (rigid) | 0.041 | 0.144 | 0.175 |
| | | $D_a$ | 1.04E-03 | 1.02E-03 | 2.18E-03 |




The extracted RDCs were more than adequate to discern the low from the high energy structures using both methods, and almost only structures representing the lowest relative energy of the ring system were present. The structural average correlated much better to experimental *J*-coupling constants and NOE correlations than did an average over all the structures. The major conformer from the RDC, NOE and *J*-coupling constant data parameters are found in Figure 8.10, where the pentacyclic system is clearly identical. The conformational populations of the pentacyclic rings are visualized in Figure 8.11 by the three dihedral angles given, as it was assumed that the indole system is rigid.

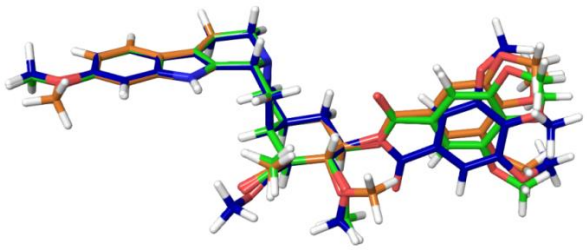


**Figure 8.10. The 3D structure which best fit the NMR data. Green: RDC data, blue: *J*-coupling constant data (RMSD = 0.024 compared to RDC) and orange NOE data (RMSD= 0.013 compared to RDC). The RMSD is based on the heavy atoms in the pentacyclic system only.**

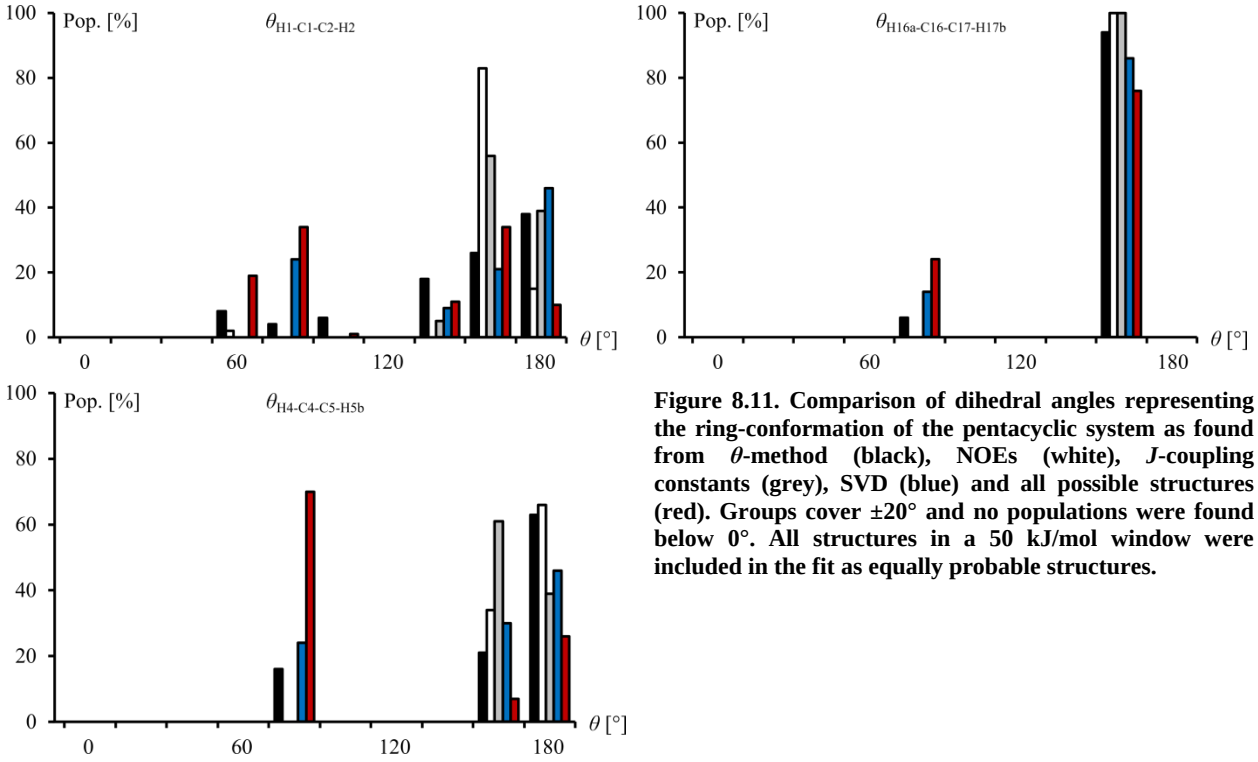



**Figure 8.11. Comparison of dihedral angles representing the ring-conformation of the pentacyclic system as found from $\theta$-method (black), NOEs (white), *J*-coupling constants (grey), SVD (blue) and all possible structures (red). Groups cover ±20° and no populations were found below 0°. All structures in a 50 kJ/mol window were included in the fit as equally probable structures.**

The $\theta$-method differentiated all diastereomers and most diastereotopic protons, other than 17 due to spectral overlap, and 21 due to similar RDCs, as seen in Table 8.7. This feat was performed including quite different conformations from all diastereomers, which could all be used to favor a wrong diastereomer. This easily demonstrates the ability to find a meaningful conformational average from multiple possible structures utilizing the $\theta$-method and solving complex 3D structural problems such as the determination of stereochemistry. The same calculations were performed using NOE and *J*-coupling constant data. The data complemented each other well, and the RDCs were just as discriminative as the other data. In the case of the diastereotopic protons at C-16, RDCs were actually needed in order to discern the two. This was caused by the overlap of the resonances of H17a and H17b which led to no discerning NOE correlations of H-16a/b, and also no *J*-coupling constants involving H-16a/b were determined. If all of the data types are utilized, the stereochemistry of all chiral and pro-chiral centers was easily solved.

**Table 8.7. RDC: *Q*-factors and ratios of *Q*-factors for wrongly assigned diastereomers of reserpine versus right assignment. Calculated by *θ* method. 17 not included due to spectral overlap of protons. Red: easily distinguished.**

| Switch # | *Q*-factor | Ratio |
|---|---|---|
| Right | 0.071 | 1.0 |
| Dia-1 | 0.630 | 8.9 |
| Dia-2 | 0.518 | 7.3 |
| Dia-3 | 0.129 | 1.8 |
| Dia-4 | 0.574 | 8.1 |
| Dia-6 | 0.709 | 10.0 |
| Dia-20 | 0.768 | 10.8 |
| Dia-5 | 0.188 | 2.6 |
| Dia-16 | 0.457 | 6.4 |
| Dia-19 | 0.108 | 1.5 |
| Dia-21 | 0.070 | 1.0 |

**Table 8.8. NOE: MAE% and ratios of MAE% for wrongly assigned diastereomers of reserpine versus right assignment from NOE (ISPA). Red: easily distinguished.**

| Switch # | MAE% | Ratio |
|---|---|---|
| Right | 5.6 | 1.0 |
| Dia-1 | 22.3 | 4.0 |
| Dia-2 | 9.3 | 1.7 |
| Dia-3 | 10.1 | 1.8 |
| Dia-4 | 12.8 | 2.3 |
| Dia-6 | 14.0 | 2.5 |
| Dia-20 | 18.5 | 3.3 |
| Dia-5 | 7.1 | 1.3 |
| Dia-16 | 7.7 | 1.4 |
| Dia-19 | 8.5 | 1.5 |
| Dia-21 | 10.2 | 1.8 |

**Table 8.9. *J*-coupling constants: MAE and ratios of MAE for wrongly assigned diastereomers of reserpine versus right assignment. Calculated by HLA. Red: easily distinguished.**

| Switch # | MAE | Ratio |
|---|---|---|
| Right | 1.0 | 1.0 |
| Dia-1 | 6.0 | 6.0 |
| Dia-2 | 2.3 | 2.3 |
| Dia-3 | 2.0 | 2.0 |
| Dia-4 | 3.8 | 3.8 |
| Dia-6 | 1.3 | 1.3 |
| Dia-20 | 4.8 | 4.8 |
| Dia-5 | 1.7 | 1.7 |
| Dia-16 | - | - |
| Dia-19 | 1.0 | 1.0 |
| Dia-21 | 2.4 | 2.4 |

**Table 8.10. Including NOEs, *J*s and RDCs in the fitting of wrongly assigned diastereomers of reserpine versus right assignment. Calculated by ISPA, HLA or *θ* method. $\Sigma=MAE_{NOE}+MAE_J/10+Q_{RDC}$. Red: easily distinguished.**

| Switch | NOE | *J* | RDC | | Ratio | | | |
|---|---|---|---|---|---|---|---|---|
| # | MAE% | MAE% | Q | Σ | NOE | *J* | RDC | Σ |
| Right | 6.0 | 15.7 | 0.112 | 0.38 | 1.0 | 1.0 | 1.0 | 1.0 |
| Dia-1 | 24.8 | 55.4 | 0.636 | 1.77 | 4.1 | 3.5 | 5.7 | 4.7 |
| Dia-2 | 9.4 | 30.4 | 0.533 | 1.03 | 1.6 | 1.9 | 4.8 | 2.7 |
| Dia-3 | 10.1 | 28.8 | 0.136 | 0.61 | 1.7 | 1.8 | 1.2 | 1.6 |
| Dia-4 | 12.8 | 49.8 | 0.594 | 1.22 | 2.1 | 3.2 | 5.3 | 3.2 |
| Dia-6 | 10.0 | 18.9 | 0.544 | 0.96 | 1.7 | 1.2 | 4.9 | 2.5 |
| Dia-20 | 18.6 | 53.5 | 0.764 | 1.68 | 3.1 | 3.4 | 6.8 | 4.4 |
| Dia-5 | 7.6 | 22.9 | 0.264 | 0.66 | 1.3 | 1.5 | 2.4 | 1.7 |
| Dia-16 | 6.6 | 15.1 | 0.545 | 0.84 | 1.1 | 1.0 | 4.9 | 2.2 |
| Dia-19 | 8.9 | 14.6 | 0.185 | 0.54 | 1.5 | 0.9 | 1.7 | 1.4 |
| Dia-21 | 18.2 | 28.0 | 0.089 | 0.75 | 3.0 | 1.8 | 0.8 | 2.0 |

The methoxy-groups could rotate and were all implicitly averaged in the fitting of the data, both using the *θ*-method and SVD fitting to a single tensor in MSPIN. Information regarding the methoxy ester could be discarded as only minor differences in back-calculated RDCs in a 3-jump model were observed due to similar angles to the magnetic field, as the ester bond was static. The phenol alcohol could be described as a two jump model. The population percentages of each position are found in Table 8.11 and are in good correlation with NOE populations.

**Table 8.11. Orientation of the C36 methyl group as determined from RDC (*θ*-method, SVD) and NOE data.**

| Conf. | *θ* method [%] | MSPIN [%] | NOE [%] |
|---|---|---|---|
| 1 | 66 | 74 | 67 |
| 2 | 34 | 26 | 33 |

A 3-jump model was assumed for the methoxy group C31 and the populations are seen in Table 8.12. While there are discrepancies between the RDC and NOE data, the RDC data consisted of a

single internuclear vector, while three NOE correlations were involved in the population determination. The NOE data was thus deemed more trustworthy.

**Table 8.12. Orientation of the C31 methyl group as determined from RDC (*θ*-method, SVD) and NOE data.**

| Conf. | *θ* method [%] | MSPIN [%] | NOE [%] |
|---|---|---|---|
| 1 | 54 | 44 | 85 |
| 2 | 8 | 10 | 10 |
| 3 | 38 | 46 | 5 |

It should be stressed that the *θ*-method, while slower for single compounds compared to SVD calculations, performed much faster when optimizing fits of multiple structures to experimental data where the SVD optimization procedure is slow. The ability to quickly screen multiple conformers and identify the correct ones is important, and while it is in principle possible to decrease calculation times for SVD methods by screening a lot of conformers in sequence, it is cumbersome, and does not really evaluate a possible fit of multiple structures. The *θ*-method may here be employed to quickly select the structures which fit experimental data.

### 8.6.3 Cinchona alkaloids

The cinchona alkaloid, named from their natural source, exhibit similar structural features with multiple rotational bonds connecting rigid parts. Possibly due to their usage in e.g. organic synthesis all structures are described quite well in literature.[111–114] Three alkaloids were investigated in order to test more flexible compounds: Quinine, cinchonine and cinchonidine. The conformational space of these compounds has been previously examined, and the literature thus provides reference studies.[111,112] The cinchona alkaloid quinine was also investigated using NOEs in Section 3.4.1.

#### Cinchonidine

The alkaloid cinchonidine is related to quinine and the structures share molecular features only differing in a methoxy group at C-8. The RDC data were fitted to multiple structures using the *θ*-method, see Table 8.13 and Table 8.14.

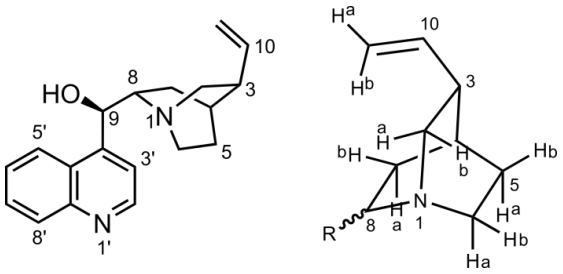



**Figure 8.12. Structure of cinchonidine and the orientation of diastereotopic protons for all cinchona alkaloids included in this thesis.**

**Table 8.13. Experimental versus back-calculated RDCs for cinchonidine. Calculated by $\theta$-method or SVD. (Polymer 8.3).**

| C | H | Exp [Hz] | $\theta$-method | SVD |
|---|---|---|---|---|
| 2 | 2a | 3.9 | 4.7 | 5.8 |
| 2 | 2b | 10.8 | 8.3 | 5.3 |
| 3 | 3 | 24.5 | 23.2 | 24.1 |
| 4 | 4 | 27.9 | 23.1 | 22.0 |
| 5 | 5a | -18.5 | -21.2 | -20.9 |
| 5 | 5b | 29.7 | 25.8 | 26.0 |
| 6 | 6a | -38.6 | -37.7 | -39.1 |
| 6 | 6b | 5.0 | 9.0 | 8.8 |
| 7 | 7a | -21.0 | -20.3 | -22.4 |
| 7 | 7b | -22.7 | -22.3 | -22.0 |
| 8 | 8 | -5.3 | -7.0 | -9.2 |
| 9 | 9 | 15.8 | 16.1 | 18.0 |
| 10 | 10 | 26.2 | 23.9 | 21.0 |
| 11 | 11a | -27.0 | -28.8 | -23.4 |
| 11 | 11b | 21.3 | 24.4 | 21.1 |
| 2' | 2' | 22.3 | 24.3 | 23.5 |
| 3' | 3' | 23.5 | 22.3 | 22.3 |
| 5' | 5' | 27.8 | 25.2 | 28.1 |
| 6' | 6' | 25.1 | 23.0 | 23.1 |
| 7' | 7' | 20.9 | 21.5 | 22.6 |
| 8' | 8' | 25.6 | 25.1 | 26.5 |
| | | *Q*-factor | 0.100 | 0.127 |
| | | $D_a$ | 8.42e-04 | 1.06e-3 |

**Table 8.14. *Q*-factors and ratios of *Q*-factors for wrongly assigned diastereomers of cinchonidine versus the right assignment. Calculated by the $\theta$-method**.

| Switch # | *Q-factor* | Ratio |
|---|---|---|
| Right | 0.100 | 1.0 |
| Dia-9 | 0.331 | 3.1 |
| Dia-7 | 0.102 | 1.0 |
| Dia-2 | 0.106 | 1.0 |
| Dia-5 | 0.372 | 3.5 |
| Dia-6 | 0.410 | 3.8 |
| Dia-11 | 0.497 | 4.6 |

The conformations of cinchonidine were previously investigated from energy computations and NMR data, primarily via *J*-coupling constants by Bürgi *et. al.*[111] and Urakawa *et al.*[117] These studies were based on a few optimized structures and their relative energies. The approach of this study thus differs, as the relative energy was disregarded and multiple conformers were used, but still the data are compared in Table 8.15. Since multiple conformers were used, and more dihedral angles are possible, the dihedral angles reported are centered on the angles determined from DFT studies. This may be viewed as the DFT structures being centroids.

**Table 8.15. Conformational populations which exhibit dihedrals $\tau$, defined as the value ±20 degrees, in percent (%). $\tau$1: C3-C4-C9-C8, $\tau$2: C4-C9-C8-N, $\tau$3: H9-C8-C9-H8. Population fitting used in MSPIN. The population of an NOE fit is used in comparison, using that of quinine, which has a similar structure as discussed in Section 3.4.1. *Other conformations identified, does not sum to 100.[a]See text. ($CDCl_3$: Polymer 8.3, DMSO-$d_6$: Polymer 8.4).**

| | $\tau$1 | | | $\tau$2 | | | | $\tau$3 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 80 | -100 | 40 | 150 | -50 | -90 | 50 | -80 | 70 | 40 | 180 | Q | $D_a$ |
| $CDCl_3$ | | | | | | | | | | | | | |
| Ref, energy(1) | 87 | 13 | - | 74 | - | - | 26 | 74 | - | | 26 | - | - |
| Ref, energy(2.1) | 80 | 11 | 9 | 69 | - | 7 | 24 | 69 | - | 7 | 24 | - | - |
| Ref, energy(2.2) | 89 | 6 | 5 | 77 | - | 12 | 11 | 77 | - | 12 | 11 | - | - |
| $\theta$ method, no scaling | 48 | 26 | 26 | 86 | - | - | 14 | 86 | - | | 14 | 0.168 | 6.48e-4 |
| $\theta$ method, scaled by 1.3 | 58 | 14 | 28 | 75 | - | - | 25 | 75 | - | | 25 | 0.100 | 8.42e-4 |
| SVD | 55 | 45 | - | 100 | - | - | - | 100 | - | - | - | 0.159 | 1.06e-3 |
| NOE (p. 54, quinine) | 67 | 12 | 21 | 67 | - | - | 33 | 67 | - | - | 33 | - | - |
| DMSO-$d_6$ | | | | | | | | | | | | | |
| Ref, energy (water)[a] | 72 | 27 | - | 45 | - | - | 54 | 45 | - | - | 54 | - | - |
| $\theta$ method, scaled by 1.2 | 39 | 46 | 15 | 28 | 20 | - | 52 | 28 | 20 | - | 52 | 0.052 | 3.96e-4 |

It is noted that the SVD method, while giving a reasonable *Q*-factor did not result in a conformational population that correlated to the rest of the NMR data. The overlay was made using the heavy atoms in the aromatic system, since it exhibited the largest absolute RDCs (after correction for the fact that the values of RDCs of parallel vectors to the magnetic field are doubled compare to perpendicular vectors). The $\theta$-method on the other hand resulted in surprisingly good populations compared to the relative energies, and compared to the NOE data of quinine, a similar structure investigated in Section 3.4.1 differing only by a methoxy group at C4', the populations fit very well between the different data types.

For comparison RDCs were also obtained from DMSO. The populations in water were used, as no data in DMSO was identified in the literature, and the energies of water and acetone are almost equal while the dielectric constant of DMSO is almost just between the two. Also the *J*-coupling constant of H11-H12 is almost equal in water and DMSO, 7.5 and 7.4 Hz respectively. Theoretically, from calculations and the *J*-coupling constant of 7.4 Hz, the ratio of $\tau$3 dihedral angle populations should now be opposite compared to chloroform populations, which was indeed observed. The *J*-coupling constants for H11 to H12 were measured to be 3.0 Hz in chloroform and 7.4 Hz in DMSO. The back-calculated *J*-coupling constants, as the average of the *J*-coupling constants of the structures determined from the RDC data by the HLA method, gave 3.2 and 5.9 Hz respectively for chloroform and DMSO in good agreement to the theoretical values.

**Cinchonine**

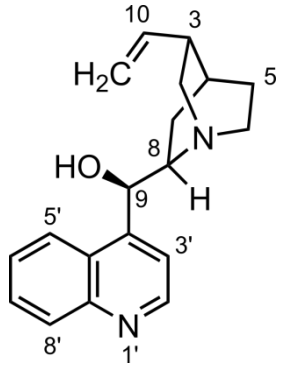
Cinchonine was compared to the work of Kowalik *et al.*[112] As for the other structures, diastereomers of internuclear vectors with numerically similar RDCs were hard to differentiate. Also differentiation of the correct assignment and the diastereomer epi-cinchonine could not initially be done based in the RDC data, Table 8.16. It should be noted that this was due to the choice of using all structures with no regard to their simulated energy, which was in this case not a preferable approach. If a cutoff was chosen at below 30 kJ/mol, it was possible to discern the stereoisomers, and it is in principle possible to solve the stereochemistry from RDCs and knowledge of relative energies between conformers.

**Table 8.16. Experimental versus back-calculated RDCs for cinchonine and the wrong diastereoisomer epi-cinchonine. Calculated by $\theta$ method, $D_a$ scaled by 1.3. *J*-coupling constants calculated using the HLA equation and the conformer populations from the RDC fit. (Polymer 8.5).**

| | | | $\theta$ method | | | |
|---|---|---|---|---|---|---|
| **C** | **H** | **Exp [Hz]** | **Cinchonine (50 kJ/mol)** | **Cinchonine (30 kJ/mol)** | **Epi-cinch. (50 kJ/mol)** | **Epi-cinch. (30 kJ/mol)** |
| 2 | 2a | -33.1 | -33.4 | -33.4 | -33.3 | -36.6 |
| 2 | 2b | 9.0 | 7.6 | 7.6 | 8.1 | 13.8 |
| 3 | 3 | -1.5 | -0.7 | -0.7 | -0.3 | 1.6 |
| 4 | 4 | 3.7 | 6.9 | 6.9 | 7.5 | 2.1 |
| 5 | 5a | -5.3 | -3.0 | -3.0 | -3.4 | -1.3 |
| 5 | 5b | -3.6 | -1.2 | -1.2 | -1.5 | -0.2 |
| 6 | 6a | 19.2 | 21.1 | 21.1 | 20.2 | 21.5 |
| 6 | 6b | 8.1 | 8.6 | 8.6 | 8.8 | 14.7 |
| 7 | 7a | -10.6 | -7.5 | -7.5 | -7.3 | -6.1 |
| 7 | 7b | -17.0 | -13.1 | -13.1 | -12.9 | -9.2 |
| 8 | 8 | 3.2 | 0.9 | 0.9 | 1.9 | 7.2 |
| 9 | 9 | 13.8 | 13.9 | 13.9 | 13.8 | 6.8 |
| 10 | 10 | -7.6 | -7.3 | -7.3 | -7.6 | -7.6 |
| 2' | 2' | 30.1 | 29.8 | 29.8 | 29.9 | 27.7 |
| 3' | 3' | 24.7 | 24.8 | 24.8 | 24.6 | 20.7 |
| 5' | 5' | 23.9 | 24.6 | 24.6 | 24.7 | 17.6 |
| 6' | 6' | 29.5 | 29.5 | 29.5 | 29.3 | 28.4 |
| 7' | 7' | 23.0 | 23.4 | 23.4 | 23.0 | 19.0 |
| 8' | 8' | 24.8 | 24.2 | 24.2 | 24.0 | 17.5 |
| | | *Q*-factor | 0.096 | 0.097 | 0.095 | 0.250 |
| | | $D_a$ | 8.30E-04 | 8.30E-04 | 8.30E-04 | 8.30E-04 |
| | | Resulting *J* | 4.8 Hz | | 2.7 Hz | |

**Table 8.17. Q-factors and ratios of *Q*-factors for wrongly assigned diastereomers of cinchonine versus right assignment. Calculated by the $\theta$ method.**

| **Switch #** | ***Q*-factor** | **Ratio** |
|---|---|---|
| Right | 0.096 | 1.0 |
| Dia-9 | 0.095 | 1.0 |
| Dia-7 | 0.098 | 1.0 |
| Dia-2 | 0.373 | 3.9 |
| Dia-6 | 0.167 | 1.7 |
| Dia-5 | 0.099 | 1.0 |




Another approach is to utilize the *J*-coupling constant between H11 and H12. For epi-cinchonine this *J*-coupling constant is reported as 9.9 Hz in chloroform compared to 4.0 Hz for cinchonine, and the diastereomers could be distinguished from the conformational differences and the associated value of the average coupling constant, where the wrong assignment (epi-cinchonine) utilized conformations which led to a markedly different *J*-coupling constant, Table 8.16.[112] Populations determined of the two diastereomers from the cinchonine RDC data is found in Table 8.18.

**Table 8.18. Conformational populations which exhibit dihedrals $\tau$, defined as the value ±20 degrees, in percent (%). $\tau$1: C3-C4-C9-O, $\tau$2: O-C9-C8-C7. Population fitting used in MSPIN. *Other conformations identified, does not sum to 100. [a]SVD fitting hold little meaning due to high $D_a$, see text. (Polymer 8.5).**

| | | $\tau$1 | | | $\tau$2 | | | |
|---|---|---|---|---|---|---|---|---|
| **Cinchonine** | **40** | **-135** | **80** | **55** | **-60** | **160** | **Q** | **$D_a$** |
| Ref, energy[112] | 95 | 5 | - | 19 | 74 | 6 | - | - |
| $\theta$ method, scaled by 1.6 | 54 | 33 | 13 | 34 | 52 | 7* | 0.096 | 8.304E-04 |
| SVD | 39 | 48 | 13 | 47 | 30 | 12* | 0.095 | 4.761E-03[a] |
| | | **$\tau$1** | | | **$\tau$2** | | | |
| **Epi-cinchonine** | **-35** | **130** | **-110** | **55** | **-65** | **175** | | |
| Ref, energy[112] | 51 | 49 | - | - | 32 | 68 | - | - |
| $\theta$ method, scaled by 1.6 | 54 | 15 | 31 | 12 | 73 | 15 | 0.095 | 8.300E-04 |

It was possible to fit the multiple conformers to the data using SVD to give a reasonable result, but there was a major concern. The alignment strength was much higher using SVD, which is rather problematic as this correlate to the individual conformers having theoretical RDCs in the hundreds

of Hz, which then averages to give reasonable back-calculated RDCs. This does not seem reasonable but the cause of this was not definitely determined. It is speculated that this was caused by the rigidity in the alignment as introduced by the MCST methodology. Since all structures are overlaid pre-fitting the conformers are not allowed to average RDCs individually and an overall average is constructed instead. In that regard it should be noted that the tensor is averaged along a fixed axis and it may well be that the axis differs between individual structures, leading to bad fit of the system. The problem was not alleviated by overlaying another part of the structure.

One could maybe limit the possible alignment strength determined by the SVD fitting, but the author has no suggestions as to how, as any limitation will result in a worse fit between experimental and back-calculated values in a natively best fit method. It was tried to utilize the alignment tensors of the individual structures as a base, by an in house written Matlab® script. Unfortunately this always resulted in the structure by which the alignment tensor was established, being by far the best fit. Also linear combinations of the alignment tensors were tried, which always resulted in worse fit compared to the native alignment tensors, and there is no theoretical evidence that this should be a viable approach.

The populations from the $\theta$ method correlated less well to the theoretical populations than those of cinchonidine, but still the fit to the theoretical populations was reasonably good, also considering the differences observed between energy calculations and NOE conformer populations for quinine.

**Quinine**

The final cinchona alkaloid to be investigated by RDCs using the $\theta$-method is quinine. The conformational space of quinine was already discussed in Section 3.4.1 obtained using NOE distances. As DFT optimized conformers of quinine were already obtained, quinine was chosen to determine the effect of having either a few, highly optimized structures (centroids) or multiple structures.

**Table 8.19. Experimental versus calculated RDCs for quinine. Calculated by $\theta$-method using structures from MM or DFT optimized structures. $D_a$ was scaled by 1.6. For SVD only DFT optimized structures were used, overlaid by the aromatic (a) or the bridged (b) part. (Polymer 8.6).**

| C | H | Exp [Hz] | $\theta$ method | | SVD | |
|---|---|---|---|---|---|---|
| | | | MM | DFT | SVD[a] | SVD[b] |
| 3 | 3 | 12.3 | 12.6 | 12.7 | 12.8 | 12.7 |
| 4 | 4 | 9.6 | 10.5 | 11.3 | 9.8 | 11.3 |
| 6 | 6a | -26.5 | -25.5 | -25.2 | -24.8 | -25.2 |
| 6 | 6b | 2.0 | 3.2 | 3.3 | 2.8 | 3.3 |
| 8 | 8 | -2.8 | -2.0 | -1.3 | -1.7 | -1.3 |
| 9 | 9 | 10.1 | 10.2 | 9.9 | 10.4 | 9.9 |
| 2' | 2' | 16.3 | 16.6 | 16.4 | 16.6 | 16.4 |
| 3' | 3' | 15.5 | 14.9 | 14.1 | 14.7 | 14.1 |
| 5' | 5' | 14.6 | 14.3 | 14.3 | 16.9 | 14.3 |
| 7' | 7' | 13.2 | 13.7 | 14.5 | 14.4 | 14.5 |
| 8' | 8' | 13.5 | 13.7 | 13.6 | 11.9 | 13.6 |
| | | Q-factor | 0.048 | 0.071 | 0.079 | 0.107 |
| | | $D_a$ | 5.54E-04 | 5.45E-04 | 4.50E-03 | 1.19E-03 |

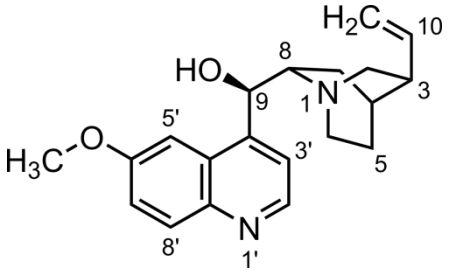



**Figure 8.13. Structure of quinine.**

The RDC data was fitted to the seven structures that were used to describe the conformational space of quinine for NOEs in Section 3.4.1.

The populations are seen in Table 8.20, comparing to theoretical conformer populations and those from the $\theta$-method and SVD.

**Table 8.20. Conformational populations which exhibit dihedrals $\tau$, defined as the value ±20 degrees, in percent (%). $\tau$1: C3-C4-C9-C8, $\tau$2: C4-C9-C8-N, $\tau$3: H9-C8-C9-H8. $D_a$ was scaled by 1.5. Population fitting used in MSPIN. [a]SVD fitting hold little meaning due to high $D_a$, see text. (Polymer 8.6).**

| | $\tau$1 | | | $\tau$2 | | | $\tau$3 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 100 | -100 | 20 | 150 | -70 | 50 | -80 | 60 | 180 | Q | $D_a$ |
| Urakawa, energy[117] | 79 | 11 | 10 | 72 | 3 | 25 | 72 | 3 | 25 | | |
| Butts, energy[IX] | 75 | 13 | 12 | 79 | 0 | 21 | 79 | 0 | 21 | | |
| $\theta$-method, no scaling | 78 | 22 | 0 | 56 | 0 | 44 | 56 | 0 | 44 | 0.138 | 3.405e-4 |
| $\theta$-method scaled by 1.6 | 59 | 31 | 0 | 31 | 7 | 62 | 31 | 7 | 62 | 0.071 | 5.448e-4 |
| SVD (arom)[a] | 34 | 51 | 15 | 29 | 37 | 34 | 29 | 37 | 34 | 0.079 | 4.50E-03 |
| SVD (bridge) | 75 | 9 | 16 | 42 | 0 | 58 | 47 | 0 | 58 | 0.107 | 1.19E-03 |
| NOE (p. 54) | 67 | 12 | 21 | 67 | 0 | 33 | 67 | 0 | 33 | | |

Analogues to the SVD fitting of cinchonine, the degree of alignment for quinine was much too high when using the aromatic system to overlay the structures, and a bit better when using the bridged system. For both the $\theta$ method and SVD, the general populations were acceptable for the rotation of the aromatic system, while the rotation around the C9-C8 bonds was not populated as expected, when compared to energy- or NOE derived populations. It was thus tried to include more structures, departing the centroids, in the approach that worked well for the cinchona alkaloids previously investigated. A reason for a worse population fit could be that the dataset was the smallest of the investigated cinchona alkaloids, due to structural overlaps and being acquired at 400 MHz, which might be problematic in the population fitting.

Entering more structures from force field calculations resulted in populations that were more alike to that of cinchonine and cinchonidine, where populations given by the rotation around C2-C9 did not correlate well to energies, but okay to populations from NOE distances, while the conformer population given by the rotation around the C8-C9 bond fitted reasonably well to the theoretical and NOE data, see Table 8.21. Thus the inclusion of multiple structures from force field simulations led to populations which correlate much better to energy- and NOE derived populations, largely comparable to the populations from the data of cinchonidine.

**Table 8.21. Conformational populations which exhibit dihedrals $\tau$, defined as the value ±20 degrees, in percent (%). $\tau$1: C3-C4-C9-C8, $\tau$2: C4-C9-C8-N, $\tau$3: H9-C8-C9-H8. *Other conformations identified, does not sum to 100. (Polymer 8.6).**

| | $\tau$1 | | | $\tau$2 | | | $\tau$3 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 90 | -100 | 20 | 150 | -80 | 50 | -80 | 60 | 180 | Q | $D_a$ |
| Urakawa, energy[117] | 79 | 11 | 10 | 72 | 3 | 25 | 72 | 3 | 25 | | |
| Butts, energy[IX] | 75 | 13 | 12 | 79 | 0 | 21 | 79 | 0 | 21 | | |
| $\theta$-method, no scaling | 43 | 16 | 41 | 91 | - | 9 | 91 | - | 9 | 0.144 | 3.46e-4 |
| $\theta$-method scaled by 1.6 | 55 | 18 | 27 | 55 | 17 | 27* | 55 | 17 | 27* | 0.048 | 5.54e-4 |
| NOE (p. 54) | 67 | 12 | 21 | 67 | - | 33 | 67 | - | 33 | - | - |

[IX] See Section 3.4.1.

It was difficult to pinpoint the rather large differences between populations identified using few and many conformers, other than the fact that multiple conformers may better illustrate the actual structural space. Still it markedly limits the usage of the $\theta$-method, and more data is needed to test reproducibility of the RDC data.

Generally the populations identified from RDC data for the cinchona alkaloids correlate nicely to energies and, especially, other NMR observables. This was actually rather surprising and the results are promising, since the cinchona alkaloids are more flexible than the structures usually investigated by RDCs, and it actually may be that RDCs hold useful non-redundant information on population of such systems.

### 8.6.4 8-phenyl-methol

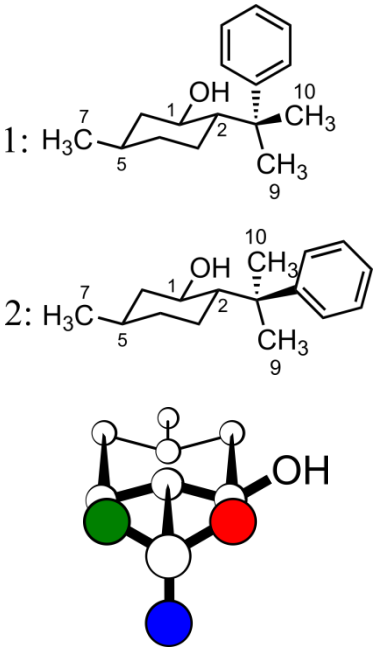
The possible pitfall of utilizing an SVD based mulita-conformer analysis, in the form of overfitting, should be addressed. For this a literature study is re-calculated by the $\theta$-method, to investigate assumptions of overfitting.

The structural conformation of 8-phenylmenthol was previously determined.[255] Here the assignment of the pro-chiral methyl groups was explored by RDCs; see Table 8.22 for the structure. Three conformers were identified by DFT calculations, which are representing a 3-jump model, i.e. conformers with the phenyl-group at all three positions indicated in Table 8.22. The RDC data were fitted to the conformers, with no assumptions of populations, in a least squares sense, from the values calculated by SVD from a common alignment tensor. The result is surprising in that the populations are almost equal and thus unaffected by the assignment of the methyl groups. This finding leads to one assignment being a better fit to the RDC data and the authors use this finding to speculate that it may be used in differentiation of the assignment. The fit between populations and RDCs seem puzzling to the current author as it is non-obvious that the population should be equal between assignments, and an identical fit should be possible from both pro-chiral assignments by inversion of the populations. It should be noted that the conclusions of the publication is not disputed, as other data is used in support.

For the investigation of the data by the $\theta$-method, the published RDC data were compared to back-calculated values obtained from the published conformers as well as eight conformations covering a 3-jump model with slight variations in the rotation. The phenyl group was treated as a 2-jump model, in practice averaging the RDCs of *ortho*- and *meta*-$^{13}$C-$^{1}$H vectors in accordance with the original publication. A scaling factor of 1.2 was used for $D_a$, positioning the ring firmly in space. The results are summarized in Table 8.22.

**Table 8.22. Experimental[255] versus calculated RDCs for 8-phenyl-methol. Calculated by $\theta$-method. The orientation of the methyl groups 9 and 10 are interchanged between stereoisomers as indicated to the right. Eight structures centered on a 3-jump model for the rotation around the C2-C8 bond or three structures previously reported were used in the fit.[255] The rotation of the C2-C8 bond is named given by the orientation of the phenyl group: Anti (blue), G+ (green) and G- (red).**

| C | H | Exp [Hz] | 8 structures [Hz] | | 3 structures [Hz][255] | |
|---|---|---|---|---|---|---|
| | | | Stereo 1 | Stereo 2 | Stereo 1 | Stereo 2 |
| 1 | 1 | 113.5 | 111.8 | 111.0 | 110.9 | 110.2 |
| 2 | 2 | 114.3 | 114.3 | 113.4 | 112.9 | 114.8 |
| 4 | 4eq | 7.7 | 5.9 | 6.6 | 8.2 | 7.4 |
| 4 | 4ax | 112.0 | 113.6 | 112.0 | 113.1 | 111.9 |
| 5 | 5 | 103.8 | 103.5 | 103.4 | 106.4 | 105.2 |
| 6 | 6eq | 8.1 | 6.0 | 7.2 | 7.5 | 7.6 |
| 6 | 6ax | 112.4 | 111.4 | 114.1 | 110.8 | 112.5 |
| 7 | 7 | 33.9 | 38.0 | 38.8 | 37.9 | 37.3 |
| 9 | 9 | -32.6 | -32.7 | -32.2 | -30.7 | -31.5 |
| 10 | 10 | -13.0 | -13.3 | -13.6 | -12.0 | -13.1 |
| *ortho* | *ortho* | 105.3 | 100.2 | 101.1 | 103.8 | 104.2 |
| *meta* | *meta* | 102.4 | 105.1 | 105.5 | 101.1 | 101.7 |
| *para* | *para* | -16.9 | -16.2 | -16.2 | -16.5 | -16.9 |
| | | Q-factor | 0.021 | 0.021 | 0.023 | 0.018 |
| | | $D_a$ | 3.76E-03 | 3.76E-03 | 3.01E-03 | 3.01E-03 |
| **Populations** | | | | | | |
| | | Anti | 0.03 | 0.00 | 0.00 | 0.00 |
| | | G+ | 0.16 | 0.62 | 0.20 | 0.62 |
| | | G- | 0.81 | 0.38 | 0.80 | 0.38 |




When fitting the conformations with the $\theta$-method excellent fits of experimental and back-calculated RDCs for both methyl assignments were determined. The assumption that the conformational population of G+ and G- could be inverted to fit the data between methyl assignments was confirmed. This makes it impossible to differentiate the two pro-chiral methyl groups, in good correlation to the experimental data. Unless one can unambiguously establish the alignment of the chair, the free rotation of the group should make the data possible for both. Also, it should be noted that since the chair only has four non-parallel vectors, the rotation of the group is used in establishing the alignment tensor, and thus relative alignment of the two groups is hardly established.

This is a prime example of how MCST fitting will at times lead to the overfitting of data. It should be noted that the authors offer another likely method of differentiation; by chemical shift calculations as the aromatic ring will shift resonances due to the inductive effect.[255] This makes the G- conformer the most likely highly populated of the G+ and G- conformers, and thus the assignment in structure 1 (Table 8.22) is the most likely assignment, in line with the conclusions reached in the original publication. Thus coupling of different data may be used to differentiate the two pro-chiral methyl groups; RDCs alone may not.

## 8.7 The Combination of Multiple Methods

The three NMR observables primarily utilized in this thesis, NOEs, $^3J$-coupling constants and RDCs, are highly compatible and orthogonal methods to gain information of 3D structures, due to differences in the mechanisms which give rise to the observables and in the averaging functions. For small molecules the three observables are seldom combined and most literature is centered on either NOEs and *J*-coupling constants or RDCs. A reason for this may be that RDCs are not yet widespread in the small molecule community and may be seen as cumbersome to obtain. Another might be that multiple programs are used for the respective calculations that do not necessarily interface well. Also the averaging of structures for RDC calculations, as well as NOEs and *J*-coupling constant calculations to a large extent, is not always straightforward if large numbers of structures are obtained.

The $\theta$-method utilizes RDC data so that it may be easily coupled to data from NOEs and *J*-coupling constants, and it is expected that this may lead to an increase of structural knowledge and ease of calculations. The three data types have been programmed to be compared to, and converge with, each other, in order to investigate 3D structures of organic compounds. This is done through the ISPA method, Karplus equations (including HLA) and the $\theta$-method. It is here utilized that all three methods are easily interfaced with each other, for single or multiple conformer systems alike.

## 8.8 Conclusion and perspectives

Population fitting of flexible compounds from RDCs still lies in the future. More data from a larger number of structures are needed to determine the proper method for structural averaging in RDC calculations. The $\theta$-method is a possible solution, where the spectroscopist or (computational) chemist is relieved from making decisions of overlays of structures and thus effectively parts of the molecule that are situated equally in space across conformers, thus easing the work. Whether the assumptions that $D_a$ may be determined from the experimental data and that a finite rotation will lead to a good representation of data holds true, may need to be finally determined from more data, but the present results are very promising. The $\theta$-method discerned stereoisomers just as well as SVD, for the relatively rigid compounds investigated. The populations from the $\theta$-method for the cinchona alkaloids correlated well to theoretical energy calculations as well as NOE populations. This is actually a little surprising but it seems to confirm an approximation made initially; that the conformer population is independent of the alignment media and $D_a$ is identical for all conformers. Whether this hold true for more compounds will need to be investigated in future work.

It is readily realized that more flexibility is needed in small molecule RDC calculations, if the technique is to find widespread usage. Otherwise the more easily accessible observables will be used instead, and for the very rigid structures that are routinely used in RDC publications, even chemical shift calculation may be an equally good and much easier alternative. This is not the case regarding flexible structures, where NOEs, *J*-coupling constants and RDCs are needed for structural information. It is thus imperative that methods are developed which may cope for structural flexibility, and this especially true for RDCs. From most data obtained thus far, which is also the data presented herein, the $\theta$-method works well, leading to meaningful conformations as well as populations, while easily handling inputs of multiple structures.

## 8.9 Experimental

**Polymers**

See Section 6.4 for information regarding synthesis of the PDMAA polymer. All other necessary information is available in figures and tables throughout this chapter.

In the table below the constituents in the synthesis of the relevant polymers are given.

| Name | Analyte (c [mM]) | Solvent | Monomer | Crosslinker (mol%) | Monomer:solvent in synthesis (v:v) | Field [MHz] |
|---|---|---|---|---|---|---|
| Polymer 8.1 | Strychnine (269.1) | $CDCl_3$ | DMAA | MBAA (0.34) | - | 800 |
| Polymer 8.2 | Reserpine (132.4) | $CDCl_3$ | DMAA | MBAA (0.17) | 1:1 $CDCl_3$ | 800 |
| Polymer 8.3 | Cinchonidine (277.1) | $CDCl_3$ | DMAA | MBAA (0.17) | 1:1 $CDCl_3$ | 800 |
| Polymer 8.4 | Cinchonidine (288.7) | DMSO-$d_6$ | DMAA | MBAA (0.16) | 5:1 $CDCl_3$ | 400 |
| Polymer 8.5 | Cinchonine (219.4) | $CDCl_3$ | DMAA | MBAA (0.17) | 1:1 $CDCl_3$ | 600 |
| Polymer 8.6 | Quinine (219.5) | $CDCl_3$ | DMAA | MBAA (0.17) | 1:1 $CDCl_3$ | 400 |

**NMR experiments**

The CLIP-HSQC experiments was used for the determination of one-bond CH coupling constants, using standard setup as described in Appendix A1. Isotropic spectra of strychnine (132 mM), reserpine (53 mM), cinchonidine (77 mM), cinchonine (86 mM) and quinine (89 mM) were acquired in 5 mm tube (0.5 mL)

**Simulations**

The modelling suite Maestro version 10.2.010 (2015) by Schrödinger was used for force field calculations,[126] using the program MacroModel version 10.8.[127,128] The MMFFs force field was used. To generate structures which should cover the conformational space of compounds, a conformational search was performed for each structure by the program MacroModel using energy cutoff of 50 kJ/mol, 100,000 steps and CPRG minimization.[127,129]

Gaussian version 09 revision B.01 was used for DFT calculations including optimizations and NMR calculations.[131] Structures were optimized to a B3LYP/6-31(d) level of theory unless otherwise stated.

# 9 Overall perspective and conclusions

Many techniques were utilized throughout this project, all focused on increasing the structural knowledge of organic and natural compounds, utilizing different parts of the NMR experimental toolbox. Large parts of the work have been focused on model compounds, usually strychnine. The true test to whether the reported methods, being NMR experiments ($S^3$ HMBCs) or RDC calculations or alignment medias ($\theta$-method and chiral media), will lead to actual useful information still lies ahead; when data of novel and/or biologically or otherwise interesting structures are investigated. The compounds of actual interest, natural products and synthetic peptides, were all investigated using NOEs and $^3J_{HH}$-coupling constants. These observables enabled the determination of stereochemistry of novel compounds that would have been difficult by use of a purely qualitative analysis, and led to novel insights into molecular structures in solution.

The first part of Chapter 3 concerns the elucidation of 3D structural features, primarily stereochemistry, of the novel natural compounds from fungal sources. While some of the structures might have been solved from qualitative data, quantitative distances added to credibility and made flexible systems solvable. Flexible used in the most rigid definition, that is. Some order is needed to gain a handle on the compounds, but the multi-structure approach, coupled to the sturdiness of the NOEs and orthogonality of the *J*-coupling constants, led to interesting results. A slight overfitting is inevitable, but the alternative of using too few structures is undesirable and limits information to a level where one could utilize only computational chemistry – and modern spectrometers and experiments are sturdy enough to avoid major errors. The natural products are followed by synthetic peptides of biological importance. Again NOEs and *J*-couplings led to structural knowledge, as either structural differences between the three HDAC inhibiting azumamides, including more knowledge of the conformational space in solution, or knowledge of the interactions between two peptides that exhibit a high degree of specific structural recognition. In the structural recognition, back-calculated distances were not the prime source of information. Instead the distances were used to limit conformational space in simulations – leading to two complimentary structures. An important lesson learned was, that while very rigid and ordered structures may be "boring" and "too easily solved", very flexible and chaotic structures are hardly solvable at all. In both cases NMR spectroscopy may help, but the approach has to be varied to fit the task at hand. The last part of the chapter focuses mostly on the biggest assumption taken, namely that the rotational correlation time is equal among all nuclei pairs. While the approaches utilized in the literature for water may work well to give an estimate to the rotational correlation time in organic solvents, care should be taken when trying to implement the rotational correlation time in calculations for organic solvent. This is largely still a work in progress.

The next chapter, Chapter 4, presents two novel NMR experimental approaches to obtain homo- or heteronuclear long-range coupling constants. These $S^3$ HMBC experiments led to easily interpretable spectra and very precise extraction of coupling constants. Especially the homonuclear variant is unrivaled in the information obtained, while the larger amount of experiments for long-range heteronuclear couplings leads to the experiment being an excellent addition to an already extensive pool. The fact that only couplings to methine carbons may be extracted is limiting, but great sensitivity, ease of use and high correlation of the extracted coupling constants to theoretical

values are significant upsides to the experiments. Both experiments are prime candidates for being used extensively in the field of carbohydrate chemistry, as methines are abundant and long-range coupling constants an already established important structural parameter, as well as more general structural elucidation.

The last three chapters (6-0) concern residual dipolar coupling constants (RDC) and their usage in structural elucidation of small molecules. The chapters may largely be seen as trying to solve the absolute and relative stereochemistry – the first for rigid structures and the latter for "flexible". Flexible is probably an overstatement, but the structures investigated were very flexible compared to the rigid compounds usually published. A considerably amount of long-range RDCs were extracted by utilizing $S^3$ HMBC *homo*, which may lead to a just as large amount of information, especially envisioned for more flexible small molecules. Enantiodiscrimination was achieved, but to a minor degree, using stretched polymers. While the prospect of utilizing stretched polymers for absolute stereochemical assignments is exciting, the current polymers do not perform to a degree where they will overtake the LC lead in the field anytime soon. But it is a start. The novel $\theta$-method led to more information on conformer populations than the routinely utilized SVD for flexible compounds, or equal amount of information in much less time for less flexible structures. While this technique is not suited for certain tasks, e.g. enantiodiscrimination, a method of gaining population information from RDCs has been searched for, since they were introduced to small molecules. While better methods may emerge, the $\theta$-method is easily implemented, relies on only the most simple of RDC formulae, and resulted in structural population information almost rivaling the NOEs. Not quite though, due to optimal averaging conditions for determining conformers of NOE distances.

All in all many challenges were faced and tried resolved. Some projects need more work, some are finished with structural knowledge gained, while other will need to stand the test of letting other users try to utilize the methods, hopefully with success. The $S^3$ HMBC experiments are "plug and play" experiments, which will hopefully lead to an immediate utilization in the small molecular NMR community. The $\theta$-method is, though promising, not expected to replace SVD and other tensor based methods due to assumptions taken up front, but may be used as a complementary method in NOE and *J*-coupling constants investigations – easily interfaced in a coupled approach to gain structural insights.

# 10 References


(1) Purcell, E. M.; Torrey, H. C.; Pound, R. V. *Phys. Rev.* **1946**, *69*, 37–38.
(2) Bloch, F.; Hansen, W. W.; Packard, M. *Phys. Rev.* **1946**, *69*, 127.
(3) Becker, E. D. *Anal. Chem.* **1993**, *65*, 295A – 302A.
(4) Rabenstein, D. L. *Anal. Chem.* **2001**, *73*, 214A – 223A.
(5) Ernst, R. R.; Anderson, W. A. *Rev. Sci. Instrum.* **1966**, *37*, 93–102.
(6) Günther, H. *NMR Spectroscopy*, 3rd ed.; Wiley-VCH: Weinheim, 2013.
(7) Neuhaus, D.; Williamson, M. *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, 1st ed.; VCH Publishers: Weinheim, 1989.
(8) Overhauser, A. W. *Phys. Rev.* **1953**, *89*, 689–700.
(9) Overhauser, A. W. *Phys. Rev.* **1953**, *92*, 411–415.
(10) Solomon, I. *Phys. Rev.* **1955**, *99*, 559–565.
(11) Levitt, M. H. *Spin Dynamics: basics of nuclear magnetic resonance*, 2nd ed.; John Wiley & Sons, Ltd: Chichester, 2008.
(12) Bax, A. D.; Davis, D. G. **1985**, *63*, 207–213.
(13) Bax, A.; Grzesiek, S. In *Encyclopedia of Magnetic Resonance, Online*; John Wiley & Sons, Ltd, 2007; pp 1–10.
(14) Hwang, T.-L.; Shaka, A. J. *J. Am. Chem. Soc.* **1992**, *114*, 3157–3159.
(15) Kjellberg, A.; Widmalm, G. *Biopolymers* **1999**, *50*, 391–399.
(16) Bross-Walch, N.; Kühn, T.; Moskau, D.; Zerbe, O. *Chem. Biodivers.* **2005**, *2*, 147–177.
(17) Macura, S.; Farmer II, B. T.; Brown, L. R. *J. Magn. Reson.* **1986**, *70*, 493–499.
(18) Andersen, N. H.; Eaton, H. L.; Lai, X. *Magn. Reson. Chem.* **1989**, *27*, 515–528.
(19) James, T. L. *Curr. Opin. Struct. Biol.* **1991**, *1*, 1042–1053.
(20) Keller, C. E.; Carper, W. R. *Magn. Reson. Chem.* **1993**, *31*, 566–572.
(21) Butts, C. P.; Jones, C. R.; Towers, E. C.; Flynn, J. L.; Appleby, L.; Barron, N. J. *Org. Biomol. Chem.* **2011**, *9*, 177–184.
(22) Borgias, B. A.; Gochin, M.; Kerwood, D. J.; James, T. L. *Prog. Nucl. Magn. Reson. Spectrosc.* **1990**, *22*, 83–100.
(23) Hu, H.; Krishnamurthy, K. *J. Magn. Reson.* **2006**, *182*, 173–177.
(24) Chi, C. N.; Strotz, D.; Riek, R.; Vögeli, B. *J. Biomol. NMR* **2015**, *62*, 63–69.
(25) Borgias, B. A.; James, T. L. *Methods Enzymol.* **1989**, *176*, 169–183.
(26) Borgias, B. A.; James, T. L. *J. Magn. Reson.* **1990**, *87*, 475–487.
(27) Post, C. B. *J. Mol. Biol.* **1992**, *224*, 1087–1101.
(28) Jones, C. R.; Butts, C. P.; Harvey, J. N. *Beilstein J. Org. Chem.* **2011**, *7*, 145–150.
(29) Butts, C. P.; Jones, C. R.; Song, Z.; Simpson, T. J. *Chem. Commun.* **2012**, *48*, 9023–9025.
(30) Butts, C. P.; Jones, C. R.; Harvey, J. N. *Chem. Commun.* **2011**, *47*, 1193–1195.
(31) Bürgi, R.; Pitera, J.; van Gunsteren, W. F. *J. Biomol. NMR* **2001**, *19*, 305–320.
(32) Daura, X.; Antes, I.; van Gunsteren, W. F.; Thiel, W.; Mark, a E. *Proteins Struct. Funct. Genet.* **1999**, *36*, 542–555.
(33) Hahn, E. L. *Phys. Rev.* **1950**, *80*, 580–594.
(34) Ramsey, N. F.; Purcell, E. M. *Phys. Rev.* **1952**, *85*, 143–144.
(35) Ramsey, N. F. *Phys. Rev.* **1953**, *91*, 303–307.
(36) Barfield, M. In *Encyclopedia of nuclear magnetic resonance*; Grant, D. M., Harris, R. K., Eds.; John Wiley & Sons, Ltd: Chichester, 1996; pp 2520–2530.
(37) Keeler, J. *Understanding NMR Spectroscopy*, 2nd ed.; John Wiley & Sons, Ltd, 2010.
(38) Engel, T. *Quantum Chemistry & Spectroscopy*, 1st ed.; Pearson Education, Inc: San Francisco, 2006.
(39) Alkorta, I.; Elguero, J. *Int. J. Mol. Sci.* **2003**, *4*, 64–92.
(40) Karplus, S.; Karplus, M. *Proc. Natl. Acad. Sci. USA* **1972**, *69*, 3204–3206.
(41) Altona, C.; Haasnoot, C. A. G. *Org. Magn. Reson.* **1980**, *13*, 417–429.
(42) Haasnoot, C. A. G.; DeLeeuw, F. A. A. M.; Altona, C. *Tetrahedron* **1979**, *36*, 2783–2792.
(43) Karplus, P. A. *Protein Sci.* **1996**, *5*, 1406–1420.
(44) Li, F.; Lee, J. H.; Grishaev, A.; Ying, J.; Bax, A. *Chemphyschem* **2015**, *16*, 572–578.
(45) Coxon, B. *Adv. Carbohydr. Chem. Biochem.* **2009**, *62*, 17–82.
(46) Bifulco, G.; Dambruoso, P.; Gomez-Paloma, L.; Riccio, R. *Chem. Rev.* **2007**, *107*, 3744–3779.
(47) Lodewyk, M. W.; Siebert, M. R.; Tantillo, D. J. *Chem. Rev.* **2012**, *112*, 1839–1862.
(48) Klages, J.; Coles, M.; Kessler, H. *Analyst* **2007**, *132*, 692–705.
(49) Jensen, F. *Introduction to Computational Chemistry*, 1st ed.; John Wiley & Sons, Ltd, 1999.
(50) Leach, A. R. *Molecular Modelling - Principles and applications*, 2nd ed.; Pearson Education Limited: Essex,

England, 2001.
(51) Gaussian DFT http://www.gaussian.com/g_tech/g_ur/k_dft.htm (accessed Nov 14, 2015).
(52) Jensen, S. R.; Flå, T.; Jonsson, D.; Monstad, S. *Phys. Chem. Chem. Phys.* **2016**.
(53) Bally, T.; Rablen, P. R. *J. Org. Chem.* **2011**, *76*, 4818–4830.
(54) Jain, R.; Bally, T.; Rablen, P. R. *J. Org. Chem.* **2009**, *74*, 4017–4023.
(55) Bagno, A.; Rastrelli, F.; Saielli, G. *Chem. Eur. J.* **2006**, *12*, 5514–5525.
(56) Chini, M. G.; Riccio, R.; Bifulco, G. *Magn. Reson. Chem.* **2008**, *46*, 962–968.
(57) Smith, S. G.; Goodman, J. M. *J. Org. Chem.* **2009**, *74*, 4597–4607.
(58) Smith, S. G.; Goodman, J. M. *J. Am. Chem. Soc.* **2010**, *132*, 12946–12959.
(59) Forsyth, D. A.; Sebag, A. B. *J. Am. Chem. Soc.* **1997**, *119*, 9483–9494.
(60) Dewick, P. M. *Medicinal Natural Products: A Biosynthetic Approach*, 3rd ed.; John Wiley & Sons, Ltd, 2009.
(61) Keller, N. P.; Turner, G.; Bennett, J. W. *Nat. Rev. Microbiol.* **2005**, *3*, 937–947.
(62) Kempken, F.; Rohlfs, M. *Fungal Ecol.* **2010**, *3*, 107–114.
(63) Calvo, A. M.; Wilson, R. A.; Bok, J. W.; Keller, N. P. *Microbiol. Mol. Biol. Rev.* **2002**, *66*, 447–459.
(64) Newman, D. J.; Cragg, G. M. *J. Nat. Prod.* **2016**, *79*, 629–661.
(65) Leeson, P. D.; Springthorpe, B. *Nat. Rev. Drug Discov.* **2007**, *6*, 881–890.
(66) Meier, J. L.; Burkart, M. D. *Chem. Soc. Rev.* **2009**, *38*, 2012–2045.
(67) Fischbach, M. A.; Walsh, C. T. *Chem. Rev.* **2006**, *106*, 3468–3496.
(68) Weber, T.; Kim, H. U. *Synth. Syst. Biotechnol.* **2016**.
(69) Varga, J.; Frisvad, J. C.; Kocsubé, S.; Brankovics, B.; Tóth, B.; Szigeti, G.; Samson, R. A. *Stud. Mycol.* **2011**, *69*, 1–17.
(70) Nielsen, K. F.; Mogensen, J. M.; Johansen, M.; Larsen, T. O.; Frisvad, J. C. *Anal. Bioanal. Chem.* **2009**, *395*, 1225–1242.
(71) Steiman, R.; Guiraud, P.; Sage, L.; Seigle-Murandi, F. *Syst. Appl. Microbiol.* **1995**, *17*, 620–624.
(72) Chiotta, M. L.; Susca, A.; Stea, G.; Mulè, G.; Perrone, G.; Logrieco, A.; Chulze, S. N. *Int. J. Food Microbiol.* **2011**, *149*, 171–176.
(73) Samson, R. A.; Noonim, P.; Meijer, M.; Houbraken, J.; Frisvad, J. C.; Varga, J. *Stud. Mycol.* **2007**, *59*, 129–145.
(74) Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry*, 6th ed.; W. H. Freeman and Company: New York, 2007.
(75) Han, Y. N.; Hwang, K. H.; Han, B. H. *Arch. Pharm. Res.* **2005**, *28*, 159–163.
(76) Hwang, K. H.; Han, Y. N.; Han, B. H. *Arch. Pharm. Res.* **2001**, *24*, 202–206.
(77) Morel, A. F.; Machado, E. C.; Wessjohann, L. A. *Phytochemistry* **1995**, *39*, 431–434.
(78) Mostardeiro, M. A.; Ilha, V.; Dahmer, J.; Caro, M. S. B.; Dalcol, I. I.; da Silva, U. F.; Morel, A. F. *J. Nat. Prod.* **2013**, *76*, 1343–1350.
(79) Park, M. H.; Suh, D.-Y.; Han, B. H. *Phytochemistry* **1996**, *43*, 701–704.
(80) Walsh, C. T.; O'Brien, R. V.; Khosla, C. *Angew. Chem. Int. Ed. Engl.* **2015**, *52*, 7098–7124.
(81) Köck, M.; Schmidt, G.; Seiple, I. B.; Baran, P. S. *J. Nat. Prod.* **2012**, *75*, 127–130.
(82) Bhushan, R.; Brückner, H. *Amino Acids* **2004**, *27*, 231–247.
(83) Marfey, P. *Carlsberg Res. Commun.* **1984**, *49*, 591–596.
(84) Wilson, W. K.; Sumpter, R. M.; Warren, J. J.; Rogers, P. S.; Ruan, B.; Schroepfer Jr., G. J. *J. Lipid Res.* **1996**, *37*, 1529–1555.
(85) Parks, L. W.; Casey, W. M. *Annu. Rev. Microbiol.* **1995**, *49*, 95–116.
(86) Zhang, Y.; Li, X.-M.; Proksch, P.; Wang, B.-G. *Steroids* **2007**, *72*, 723–727.
(87) Petersen, L. M.; Hoeck, C.; Frisvad, J. C.; Gotfredsen, C. H.; Larsen, T. O. *Molecules* **2014**, *19*, 10898–10921.
(88) TePaske, M. R.; Gloer, J. B. *Tetrahedron* **1989**, *45*, 4961–4968.
(89) Gloer, J. B.; Tepaske, M. R.; Sima, J. S.; Wicklow, D. T.; Dowd, P. F. *J. Org. Chem.* **1988**, *53*, 5457–5460.
(90) Nakao, Y.; Yoshida, S.; Matsunaga, S.; Shindoh, N.; Terada, Y.; Nagai, K.; Yamashita, J. K.; Ganesan, A.; Van Soest, R. W. M.; Fusetani, N. *Angew. Chemie* **2006**, *118*, 7715–7719.
(91) Bertrand, P. *Eur. J. Med. Chem.* **2010**, *45*, 2095–2116.
(92) de Ruijter, A. J. M.; van Gennip, A. H.; Caron, H. N.; Kemp, S.; van Kuilenburg, A. B. P. *Biochem. J.* **2003**, *370*, 737–749.
(93) Maolanon, A. R.; Villadsen, J. S.; Christensen, N. J.; Hoeck, C.; Friis, T.; Harris, P.; Gotfredsen, C. H.; Fristrup, P.; Olsen, C. A. *J. Med. Chem.* **2014**, *57*, 9644–9657.
(94) Izzo, I.; Maulucci, N.; Bifulco, G.; De Riccardis, F. *Angew. Chemie* **2006**, *118*, 7719–7722.
(95) Maulucci, N.; Chini, M. G.; Micco, S. Di; Izzo, I.; Cafaro, E.; Russo, A.; Gallinari, P.; Paolini, C.; Nardi, M. C.; Casapullo, A.; Riccio, R.; Bifulco, G.; Riccardis, F. De. *J. Am. Chem. Soc.* **2007**, *129*, 3007–3012.
(96) Vögeli, B.; Güntert, P.; Riek, R. *Mol. Phys.* **2013**, *111*, 437–454.
(97) Levy, Y.; Onuchic, N. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 389–415.

(98) Crabtree, G. R.; Clipstone, N. A. *Annu. Rev. Biochem.* **1994**, *63*, 1045–1083.
(99) Sundberg, E. J.; Mariuzza, R. A. *Adv. Protein Chem.* **2002**, *61*, 119–160.
(100) Babine, R. E.; Bender, S. L. *Chem. Rev.* **1997**, *97*, 1359–1472.
(101) Renil, M.; Ferreras, M.; Delaisse, J. M.; Foged, N. T.; Meldal, M. *J. Pept. Sci.* **1998**, *4*, 195–210.
(102) Meldal, M.; Wu, B.; Diness, F.; Michael, R.; Hagel, G. *ChemBioChem* **2011**, *12*, 2463–2470.
(103) Diamandis, E. P.; Christopoulos, T. K. *Clin. Chem.* **1991**, *37*, 625–636.
(104) Li, M.; Hoeck, C.; Schoffelen, S.; Gotfredsen, C. H.; Meldal, M. *Chem. Eur. J.* **2016**, *22*, 7206–7214.
(105) Sigurskjolds, B. W.; Bundle, D. R. *J. Biol. Chem.* **1992**, *267*, 8371–8376.
(106) Kaur, H.; Datt, M.; Ekka, M. K.; Mittal, M.; Singh, A. K.; Kumaran, S. *Biochimie* **2011**, *93*, 175–186.
(107) Sasaki, S.; Takagi, M.; Tanaka, Y.; Maeda, M. *Tetrahedron Lett.* **1996**, *37*, 85–88.
(108) Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 808–813.
(109) Thordarson, P. *Chem. Soc. Rev.* **2011**, *40*, 1305–1323.
(110) Evans, J. N. S. *Biomolecular NMR spectroscopy*, 1st ed.; Oxford University Press: Oxford, 1995.
(111) Burgi, T.; Baiker, A. *J. Am. Chem. Soc.* **1998**, *120*, 12920–12926.
(112) Kowalik, J. T.; Lipinska, T.; Oleksyn, B. J.; Sliwinski, J. *Enantiomer* **1999**, *4*, 389–410.
(113) Dijkstra, G. D. H.; Kellogg, R. M.; Wynberg, H. *J. Org. Chem.* **1990**, *55*, 6121–6131.
(114) Yeboah, E. M. O.; Yeboah, S. O.; Singh, G. S. *Tetrahedron* **2011**, *67*, 1725–1762.
(115) Silva, T. H. A.; Oliveira, A. B.; Almeida, W. B. De. *Struct. Chem.* **1997**, *8*, 95–107.
(116) Sen, A.; Bouchet, A.; Lepe, V.; Barbu-debus, K. Le; Scuderi, D.; Piuzzi, F.; Zehnacker-Rentien, A. *J. Phys. Chem. A* **2012**, *116*, 8334–8344.
(117) Urakawa, A.; Meier, D. M.; Rüegger, H.; Baiker, A. *J. Phys. Chem. A* **2008**, *112*, 7250–7255.
(118) Efimov, S. V.; Khodov, I. A.; Ratkova, E. L.; Kiselev, M. G.; Berger, S.; Klochkov, V. V. *J. Mol. Struct.* **2016**, *1104*, 63–69.
(119) Pendrill, R.; Säwén, E.; Widmalm, G. *J. Phys. Chem. B* **2013**, *117*, 14709–14722.
(120) Kessler, H.; Griesinger, C.; Kerssebaum, R.; Wagner, K.; Ernst, R. R. *J. Am. Chem. Soc.* **1987**, *109*, 607–609.
(121) Mackeen, M.; Almond, A.; Cumpstey, I.; Enis, S. C.; Kupce, E.; Butters, T. D.; Fairbanks, A. J.; Dwek, R. A.; Wormald, M. R. *Org. Biomol. Chem.* **2006**, *4*, 2241–2246.
(122) Mao, X.; Zhang, T.; Baur, M.; Kessler, H. *J. Chem. Phys.* **1999**, *111*, 8253–8254.
(123) Lee, D.; Hilty, C.; Wider, G.; Wüthrich, K. *J. Magn. Reson.* **2006**, *178*, 72–76.
(124) Hwang, T.-L.; Kadkhodaei, M.; Mohebbi, A.; Shaka, A. J. *Magn. Reson. Chem.* **1992**, *30*, S24–S34.
(125) Olsson, U.; Säwën, E.; Stenutz, R.; Widmalm, G. *Chem. Eur. J.* **2009**, *15*, 8886–8894.
(126) Maestro. Schrödinger, LLC: New York 2015.
(127) Macromodel. Schrödinger, LLC: New York 2015.
(128) Desmond. D. E. Shaw Research: New York 2015.
(129) Chang, G.; Guida, W. C.; Still, W. C. *J. Am. Chem. Soc.* **1989**, *111*, 4379–4386.
(130) Chen, I.; Foloppe, N. *Bioorg. Med. Chem.* **2013**, *21*, 7898–7920.
(131) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. S.; M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. M.; G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. H.; A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. H.; M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. N.; Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J.; J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. B.; K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. N.; K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. T.; M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. C.; V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. S.; O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. O.; R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. V.; P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. D.; O. Farkas, J. B. Foresman, J. V. Ortiz, J. C.; D. J. Fox. Gaussian, Inc., Wallingford CT 2010.
(132) Kjærulff, L. NMR structural studies of oligosaccharides and other natural products, Technical University of Denmark, 2014.
(133) Gattin, Z.; Zaugg, J.; van Gunsteren, W. F. *Chemphyschem* **2010**, *11*, 830–835.
(134) Wu, A.; Cremer, D. *Int. J. Mol. Sci.* **2003**, *4*, 158–192.
(135) Rundlöf, T.; Kjellberg, A.; Damberg, C.; Nishida, T.; Widmalm, G. *Magn. Reson. Chem.* **1998**, *36*, 839–847.
(136) Hansen, P. E. *Prog. Nucl. Magn. Reson. Spectrosc.* **1981**, *14*, 175–296.
(137) Tvaroska, I.; Hricovíni, M.; Petráková, E. *Carbohydr. Res.* **1989**, *189*, 359–362.
(138) Barfield, M. *J. Chem. Phys.* **1964**, *41*, 3825–3832.
(139) Barfield, M. *J. Am. Chem. Soc.* **1971**, *93*, 1066–1071.
(140) Cen-Pacheco, F.; Rodríguez, J.; Norte, M.; Fernández, J. J.; Hernández Daranas, A. *Chem. Eur. J.* **2013**, *19*, 8525–8532.
(141) García, P.; Martín-Pastor, M.; De Lera, Á. R.; Álvarez, R. *Magn. Reson. Chem.* **2010**, *48*, 543–549.
(142) Schmidt, J. M.; Sørensen, O. W.; Ernst, R. R. *J. Magn. Reson. Ser. A* **1994**, *109*, 80–89.

(143) Mahi, L.; Duplan, J. C. *Magn. Reson. Chem.* **1997**, *35*, 379–383.
(144) Barfield, M.; Chakrabarti, B. *Chem. Rev.* **1969**, *69*, 757–778.
(145) Parella, T.; Espinosa, J. F. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 17–55.
(146) Marquez, B. L.; Gerwick, W. H.; Williamson, R. T. *Magn. Reson. Chem.* **2001**, *39*, 499–530.
(147) Katritzky, A. R.; Akhmedov, N. G.; Güven, A.; Doskocz, J.; Akhmedova, R. G.; Majumder, S.; Hall, C. D. *J. Mol. Struct.* **2006**, *787*, 131–147.
(148) Elgamal, M. H. A.; Elewa, N. H.; Elkhrisy, E. A. M.; Duddeck, H. *Phytochemistry* **1978**, *18*, 139–143.
(149) Hansen, M.; Jakobsen, H. J. *J. Magn. Reson.* **1975**, *20*, 520–529.
(150) Zhu, G.; Renwick, A.; Bax, A. *J. Magn. Reson. Ser. A* **1994**, *110*, 257–261.
(151) Zhu, G.; Bax, A. *J. Magn. Reson. Ser. A* **1993**, *104*, 353–357.
(152) Krishnamurthy, V. V. *J. Magn. Reson. Ser. A* **1996**, *121*, 33–41.
(153) Butts, C. P.; Heise, B.; Tatolo, G. *Org. Lett.* **2012**, *14*, 3256–3259.
(154) Meissner, A.; Sørensen, O. *Magn. Reson. Chem.* **2001**, *39*, 49–52.
(155) Kozminski, W.; Nanz, D. *J. Magn. Reson.* **2000**, *142*, 294–299.
(156) Saurí, J.; Nolis, P.; Castañar, L.; Virgili, A.; Parella, T. *J. Magn. Reson.* **2012**, *224*, 101–106.
(157) Kövér, K. E.; Prakash, O.; Hruby, V. J. *J. Magn. Reson. Ser. A* **1993**, *103*, 92–96.
(158) Kövér, K. E.; Hruby, V. J.; Uhrín, D. *J. Magn. Reson.* **1997**, *129*, 125–129.
(159) Kozminski, W. *J. Magn. Reson.* **1999**, *137*, 408–412.
(160) Gil, S.; Espinosa, J. F.; Parella, T. *J. Magn. Reson.* **2010**, *207*, 312–321.
(161) Gil, S.; Espinosa, J. F.; Parella, T. *J. Magn. Reson.* **2011**, *213*, 145–150.
(162) Ding, K. *Magn. Reson. Chem.* **2000**, *38*, 321–323.
(163) Saurí, J.; Parella, T. *Magn. Reson. Chem.* **2012**, *50*, 717–721.
(164) Kurz, M.; Schmieder, P.; Kessler, H. *Angew. Chem. Int. Ed. Engl* **1991**, *30*, 1329–1331.
(165) Sørensen, M. D.; Meissner, A.; Sørensen, O. W. *J. Magn. Reson.* **1999**, *137*, 237–242.
(166) Kjaerulff, L.; Benie, A. J.; Hoeck, C.; Gotfredsen, C. H.; Sørensen, O. W. *J. Magn. Reson.* **2016**, *263*, 101–107.
(167) Nyberg, N. T.; Duus, J. Ø.; Sørensen, O. W. *J. Am. Chem. Soc.* **2005**, *127*, 6154–6155.
(168) Nyberg, N. T.; Duus, J. Ø.; Sørensen, O. W. *Magn. Reson. Chem.* **2005**, *43*, 971–974.
(169) Benie, A. J.; Sørensen, O. W. *Magn. Reson. Chem.* **2006**, *44*, 739–743.
(170) Benie, A. J.; Sørensen, O. W. *J. Magn. Reson.* **2007**, *184*, 315–321.
(171) Meissner, A.; Duus, J. Ø.; Sørensen, O. W. *J. Biomol. NMR* **1997**, *10*, 89–94.
(172) Sørensen, M. D.; Meissner, A.; Sørensen, O. W. *J. Biomol. NMR* **1997**, *10*, 181–186.
(173) Meissner, A.; Duus, J. Ø.; Sørensen, O. W. *J. Magn. Reson.* **1997**, *128*, 92–97.
(174) Meissner, A.; Schulte-herbrüggen, T.; Briand, J.; Meissner, A.; Schulte-herbruggen, T.; Srensen, O. W. *Mol. Phys.* **1998**, *95*, 1137–1142.
(175) Sørensen, O. W. *Prog. Nucl. Magn. Reson. Spectrosc.* **1989**, *21*, 503–569.
(176) Crecely, K. M.; Crecely, R. W.; Goldstein, J. H. *J. Mol. Spectrosc.* **1971**, *37*, 252–259.
(177) Meissner, A.; Schulte-Herbrüggen, T.; Sørensen, O. W. *J. Am. Chem. Soc.* **1998**, *120*, 7989–7990.
(178) Wiitala, K. W.; Cramer, C. J.; Hoye, T. R. *Magn. Reson. Chem.* **2007**, *45*, 819–829.
(179) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664–675.
(180) Schmidt, M.; Reinscheid, F.; Sun, H.; Abromeit, H.; Scriba, G. K. E.; Sönnichsen, F. D.; John, M.; Reinscheid, U. M. *Eur. J. Org. Chem.* **2014**, 1147–1150.
(181) Edden, R. A. E.; Keeler, J. *J. Magn. Reson.* **2004**, *166*, 53–68.
(182) Blechta, V.; del Río-Portilla, F.; Freeman, R. *Magn. Reson. Chem.* **1994**, *32*, 134–137.
(183) Hatcher, E.; Säwén, E.; Widmalm, G.; Mackerell, Jr., A. D. *J. Phys. Chem. B* **2011**, *115*, 597–608.
(184) Martin-Pastor, M.; Bush, C. A. *J. Biomol. NMR* **2001**, *19*, 125–139.
(185) Hogben, H. J.; Krzystyniak, M.; Charnock, G. T. P.; Hore, P. J.; Kuprov, I. *J. Magn. Reson.* **2011**, *208*, 179–194.
(186) Sarfati, M.; Lesot, P.; Merlet, D.; Courtieu, J. *Chem. Commun.* **2000**, 2069–2081.
(187) Thiele, C. M.; Berger, S. *Org. Lett.* **2003**, *5*, 705–708.
(188) Thiele, C. M. *J. Org. Chem.* **2004**, *69*, 7403–7413.
(189) Meyer, N. C.; Krupp, A.; Schmidts, V.; Thiele, C. M.; Reggelin, M. *Angew. Chemie Int. Ed.* **2012**, *51*, 8334–8338.
(190) Marx, A.; Schmidts, V.; Thiele, C. M. *Magn. Reson. Chem.* **2009**, *47*, 734–740.
(191) Montag, T.; Thiele, C. M. *Chem. Eur. J.* **2013**, *19*, 2271–2274.
(192) Krupp, A.; Reggelin, M. *Magn. Reson. Chem.* **2012**, *50*, S45–S52.
(193) Luy, B.; Kobzar, K.; Kessler, H. *Angew. Chemie Int. Ed.* **2004**, *43*, 1092–1094.
(194) Gayathri, C.; Tsarevsky, N. V.; Gil, R. R. *Chem. Eur. J.* **2010**, *16*, 3622–3626.
(195) Thiele, C. M. *Eur. J. Org. Chem.* **2008**, 5673–5685.

(196) Kummerlöwe, G.; Luy, B. *Trends Anal. Chem.* **2009**, *28*, 483–493.
(197) Thiele, C. M. *Concepts Magn. Reson. Part A* **2007**, *30A*, 65–80.
(198) Marx, A.; Thiele, C. *Chem. Eur. J.* **2009**, *15*, 254–260.
(199) Kummerlöwe, G.; McCord, E. F.; Cheatham, S. F.; Niss, S.; Schnell, R. W.; Luy, B. *Chem. Eur. J.* **2010**, *16*, 7087–7089.
(200) Freudenberger, J. C.; Knör, S.; Kobzar, K.; Heckmann, D.; Paululat, T.; Kessler, H.; Luy, B. *Angew. Chemie Int. Ed.* **2005**, *44*, 423–426.
(201) Blackledge, M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2005**, *46*, 23–61.
(202) Losonczi, J. A.; Andrec, M.; Fischer, M. W. F.; Prestegard, J. H. *J. Magn. Reson.* **1999**, *138*, 334–342.
(203) Zweckstetter, M. *Nat. Protoc.* **2008**, *3*, 679–690.
(204) Navarro-Vázquez, A. *Magn. Reson. Chem.* **2012**, *50*, S73–S79.
(205) Tzvetkova, P.; Luy, B.; Simova, S. *Magn. Reson. Chem.* **2012**, *50*, S92–S101.
(206) Enthart, A.; Freudenberger, J. C.; Furrer, J.; Kessler, H.; Luy, B. *J. Magn. Reson.* **2008**, *192*, 314–322.
(207) Timári, I.; Kaltschnee, L.; Kolmer, A.; Adams, R. W.; Nilsson, M.; Thiele, C. M.; Morris, G. A.; Kövér, K. E. *J. Magn. Reson.* **2014**, *239*, 130–138.
(208) Snider, J. D.; Troche-Pesqueira, E.; Woodruff, S. R.; Gayathri, C.; Tsarevsky, N. V.; Gil, R. R. *Magn. Reson. Chem.* **2012**, *50*, S86–S91.
(209) Bax, A.; Grishaev, A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 563–570.
(210) Schmidt, M.; Sun, H.; Leonov, A.; Griesinger, C.; Reinscheid, U. M. *Magn. Reson. Chem.* **2012**, *50*, S38–S44.
(211) Lesot, P.; Sarfati, M.; Courtieu, J. *Chem. Eur. J.* **2003**, *9*, 1724–1745.
(212) Arnold, L.; Marx, A.; Thiele, C. M.; Reggelin, M. *Chem. Eur. J.* **2010**, *16*, 10342–10346.
(213) Lei, X.; Xu, Z.; Sun, H.; Wang, S.; Griesinger, C.; Peng, L.; Gao, C.; Tan, R. X. *J. Am. Chem. Soc.* **2014**, *136*, 11280–11283.
(214) Zong, W.; Li, G.; Cao, J.; Lei, X.; Hu, M.; Sun, H.; Griesinger, C.; Tan, R. X. *Angew. Chemie Int. Ed.* **2016**, *55*, 3690–3693.
(215) Luy, B.; Kobzar, K.; Knör, S.; Furrer, J.; Heckmann, D.; Kessler, H. *J. Am. Chem. Soc.* **2005**, *127*, 6459–6465.
(216) Freudenberger, J. C.; Spiteller, P.; Bauer, R.; Kessler, H.; Luy, B. *J. Am. Chem. Soc.* **2004**, *126*, 14690–14691.
(217) Haberz, P.; Farjon, J.; Griesinger, C. *Angew. Chemie - Int. Ed.* **2005**, *44*, 427–429.
(218) Meier, S.; Häussinger, D.; Grzesiek, S. *J. Biomol. NMR* **2002**, *24*, 351–356.
(219) Kobzar, K.; Kessler, H.; Luy, B. *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 3145–3147.
(220) Marx, A.; Böttcher, B.; Thiele, C. M. *Chem. Eur. J.* **2010**, *16*, 1656–1663.
(221) Kuchel, P. W.; Chapman, B. E.; Müller, N.; Bubb, W. A.; Philp, D. J.; Torres, A. M. *J. Magn. Reson.* **2006**, *180*, 256–265.
(222) Kummerlöwe, G.; Halbach, F.; Laufer, B.; Luy, B. *Open Spectrosc. J.* **2008**, *2*, 29–33.
(223) Kummerlöwe, G.; Kiran, M. U.; Luy, B. *Chem. Eur. J.* **2009**, *15*, 12192–12195.
(224) Luy, B. *J. Indian Inst. Sci.* **2010**, *90*, 119–132.
(225) Sass, J.; Cordier, F.; Hoffmann, A.; Cousin, A.; Omichinski, J. G.; Löwen, H.; Grzesiek, S. *J. Am. Chem. Soc.* **1999**, *121*, 2047–2055.
(226) Thiele, C. M.; Maliniak, A.; Stevensson, B. *J. Am. Chem. Soc.* **2009**, *131*, 12878–12879.
(227) Tian, F.; Bolon, P. J.; Prestegard, J. H. *J. Am. Chem. Soc.* **1999**, *121*, 7712–7713.
(228) Yan, J.; Delaglio, F.; Kaerner, A.; Kline, A. D.; Mo, H.; Shapiro, M. J.; Smitka, T. a; Stephenson, G. a; Zartler, E. R. *J. Am. Chem. Soc.* **2004**, *126*, 5008–5017.
(229) Möglich, A.; Wenzler, M.; Kramer, F.; Glaser, S. J.; Brunner, E. *J. Biomol. NMR* **2002**, *23*, 211–219.
(230) Wainer, I. W. *Hum. Psychopharmacol. Clin. Exp.* **2001**, *16*, S73–S77.
(231) Maier, N. M.; Franco, P.; Lindner, W. *J. Chromatogr. A* **2001**, *906*, 3–33.
(232) Mcconnell, O.; Ii, A. B.; Balibar, C.; Byrne, N.; Cai, Y.; Carter, G. U. Y.; Chlenov, M.; Di, L. I.; Fan, K.; Goljer, I.; He, Y.; Herold, D. O. N.; Kagan, M.; Kerns, E.; Koehn, F.; Kraml, C.; Marathias, V.; Marquez, B.; Mcdonald, L.; Nogle, L.; Petucci, C.; Schlingmann, G.; Tawa, G.; Tischler, M.; Williamson, R. T.; Sutherland, A.; Watts, W.; Young, M.; Zhang, M.; Zhang, Y.; Zhou, D.; Ho, D. *Chirality* **2007**, *19*, 658–682.
(233) Clayden, J.; Greeves, N.; Warren, S.; Wothers, P. *Organic Chemistry*; OUP Oxford, 2001.
(234) Wenzel, T. J.; Chisholm, C. D. *Prog. Nucl. Magn. Reson. Spectrosc.* **2011**, *59*, 1–63.
(235) Wenzel, T. J.; Chisholm, C. D. *Chirality* **2011**, *23*, 190–214.
(236) Breton, R. C.; Reynolds, W. F. *Nat. Prod. Rep.* **2013**, *30*, 501–524.
(237) Parker, D. *Chem. Rev.* **1991**, *91*, 1441–1457.
(238) Balzano, F.; Brotin, T.; Bredehöft, J. H.; Dutasta, J.-P.; Evans, A. C.; Giri, C.; Guy, L.; Hoffmann, S. V.; Jones, N. C.; Manoli, K.; Magliulo, M.; Martinez, A.; Meinert, C.; Meierhenrich, U. J.; Sorochinsky, A. E.; Soloshonok, V. A.; Torsi, L.; Trapp, O.; Uccello-Barretta, G.; Wenzel, T. J. *Differentiation of Enantiomers II*, 1st ed.; Schurig, V., Ed.; Springer: Heidelberg, 2013.

(239) Yuki, H.; Ohta, K.; Uno, K.; Murahashi, S. *J. Polym. Sci. A* **1968**, *6*, 829–841.
(240) Kitayama, T.; Shibuya, W.; Katsukawa, K. *Polym. J.* **2002**, *34*, 405–409.
(241) Porter, N. A.; Allen, T. R.; Breyer, R. A. *J. Am. Chem. Soc.* **1992**, *114*, 7676–7683.
(242) Porter, N. A.; Rosenstein, I. J.; Breyer, R. A.; Bruhnke, J. D.; Wu, W. X.; McPhail, A. T. *J. Am. Chem. Soc.* **1992**, *114*, 7664–7676.
(243) Hatada, K.; Kitayama, T.; Ute, K. *Prog. Polym. Sci.* **1988**, *13*, 189–276.
(244) Berger, R.; Courtieu, J.; Gil, R. R.; Griesinger, C.; Köck, M.; Lesot, P.; Luy, B.; Merlet, D.; Navarro-Vázquez, A.; Reggelin, M.; Reinscheid, U. M.; Thiele, C. M.; Zweckstetter, M. *Angew. Chemie Int. Ed.* **2012**, *51*, 8388–8391.
(245) Frank, A. O.; Freudenberger, J. C.; Shaytan, A. K.; Kessler, H.; Luy, B. *Magn. Reson. Chem.* **2015**, *53*, 213–217.
(246) Canales, A.; Jiménez-Barbero, J.; Martín-Pastor, M. *Magn. Reson. Chem.* **2012**, *50*, S80–S85.
(247) Matlab documentation - SVD http://se.mathworks.com/help/matlab/ref/svd.html (accessed Apr 15, 2016).
(248) Kolmer, A.; Edwards, L. J.; Kuprov, I.; Thiele, C. M. *J. Magn. Reson.* **2015**, *261*, 101–109.
(249) Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1999**, *121*, 4690–4695.
(250) Verdier, L.; Sakhaii, P.; Zweckstetter, M.; Griesinger, C. *J. Magn. Reson.* **2003**, *163*, 353–359.
(251) Montalvao, R.; Camilloni, C.; De Simone, A.; Vendruscolo, M. *J. Biomol. NMR* **2014**, *58*, 233–238.
(252) Camilloni, C.; Vendruscolo, M. *J. Phys. Chem. B* **2015**, *119*, 653–661.
(253) Clore, G. M.; Gronenborn, A. M.; Tjandra, N. *J. Magn. Reson.* **1998**, *131*, 159–162.
(254) Bifulco, G.; Riccio, R.; Martin, G. E.; Buevich, A. V; Williamson, R. T. *Org. Lett.* **2013**, *15*, 654–657.
(255) Sánchez-Pedregal, V. M.; Santamaría-Fernández, R.; Navarro-Vázquez, A. *Org. Lett.* **2009**, *11*, 1471–1474.

# Appendices

## Contents

# A1. NMR

## A1.1. Spectrometers

**DTU**

All spectra were acquired at one of the five NMR spectrometers given below. Spectra were acquired using standard pulse sequences at 25 °C unless stated otherwise. The spectra were recorded using 5 mm NMR tubes when possible (solvent volume 500 µL) or 3 mm tubes (solvent volume 180 µL). Chemical shifts are always given in ppm, and coupling constants in Hz.

- Varian Inova 500 MHz with a 5 mm HCP probe (499.87 MHz for $^{1}$H, 125.70 MHz for $^{13}$C).
- Bruker Avance III 400 MHz with a 5 mm BBO smartprobe (400.23 MHz for $^{1}$H, 100.64 MHz for $^{13}$C).
- Bruker Ascend 400 MHz with a 5 mm H-br.band dual channel z-gradient Prodigy cryoprobe (400.13 MHz for $^{1}$H, 100.61 MHz for $^{13}$C).
- Bruker DRX 600 MHz with a 5 mm BBO smartprobe (600.13 MHz for $^{1}$H, 150.92 MHz for $^{13}$C).
- Bruker Avance 800 MHz with a 5 mm TCI cryoprobe (798.80 MHz for $^{1}$H, 200.86 MHz for $^{13}$C).

**External stay**

The NMR spectrometers used for acquisitions at the University of Bristol are given below.

- Varian VNMRS DirectDrive 500 MHz with a broadband two-channel OneNMR probe (499.66 MHz for $^{1}$H, 125.64 MHz for $^{13}$C)
- Varian VNMRS DirectDrive 600 MHz spectrometer equipped with an indirect observe cryoprobe (600.05 MHz for $^{1}$H, 150.88 MHz for $^{13}$C).

## A1.2. Solvents

The solvents used are given below along with the solvent resonances used for referencing the spectra.[1] Chemical shifts ($\delta$) are in ppm and scalar couplings are reported in hertz (Hz).

| Solvent | $\delta_H$ [ppm] | $\delta_C$ [ppm] |
|---|---|---|
| DMSO-$d_6$ | 2.50 | 39.5 |
| $CDCl_3$ | 7.26 | 77.2 |
| $CD_3OD$ | 3.31 | 49.0 |
| $D_2O$ | 4.79 | - |

## A1.3. Acquisition

Examples of the basic setup of the most used literature experiments are found below. The pulse sequence given are examples of sequences often used on Bruker spectrometers.

**1D $^{1}$H** experiments (w/wo water suppression) were used to acquire 1D proton spectra. A standard acquisition was as follows: si (Fourier transform size) = 64k, rd (relaxation delay) = 1.0 s, ns (number of scans) = 32, ds (number of dummy scans)= 8. (**zg30**)

**CLIP-HSQC** experiment was used to acquire f2-coupled HSQC spectra for the extraction of $^{1}J_{CH}$ and $^{1}T_{CH}$ (RDCs). A standard acquisition was as follows: si = 4096 , ni (number of increments) = 256, rd = 1.0 s, ns = 8, ds = 16. The spectra were zero filled to 16k in the F2 and 1k in the F1 dimension. Coupling constants were extracted by overlay of 1D slices through F1. (**CLIP_hsqcetgp**, pulse sequence acquired from http://www.ioc.kit.edu/luy/110.php website of Prof. Dr. Burkhard Luy, Institut für Organische Chemie, Karlsruher Institut für Technologie)[2]

**NOESY** experiments (w/wo water suppression) were used to acquire 2D NOESY spectra. A standard acquisition was as follows: si = 4096 , ni = 256, ns = 8, ds = 32. The mixing time was usually 50-300 ms and the relaxation delay depends on the $T_1$ of the protons (usually around 1-5 s used). The spectra were zero filled to 8k in the F2 and 1k in the F1 dimension. Integrals were determined from the spectra in Topspin. (**noesyph, noesyphpr, noesygpph19**)[3–6]

**ROESY** experiments (w/wo water suppression) were used to acquire 2D T-ROESY spectra. A standard acquisition was as follows: si = 4096 , ni = 256, ns = 8, ds = 32. The mixing time was usually 50-300 ms and the relaxation delay depends on the $T_1$ of the protons (usually around 1-5 s used). The spectra were zero filled to 8k in the F2 and 1k in the F1 dimension. Integrals were determined from the spectra in Topspin. (**roesyph.2, roesyphpr.2, roesygpph19.2**)[7,8,5,6]

**DPFGSE NOESY** (double-pulsed-field-gradient-spin-echo) or **SPFGSE NOESY** (single-pulsed-field-gradient-spin-echo) experiment was used for 1D NOESY experiments. A standard acquisition was as follows: si = 64k, rd = 1.0 s, ns = 128, ds = 8. The mixing time was usually 300-500 ms. The spectra were zero filled to 128k. Integrals were determined by setting the integral of the irradiated resonance equal to -1,000 and determining the other relative integrals (PANIC approach).

**DQF-COSY** experiments were used to acquire 2D COSY spectra. A standard acquisition was as follows: si = 4096 or 8096 , ni = 512, ns = 8, ds = 32, rd = 1.0 s. The spectra were zero filled to 8k in the F2 and 1k in the F1 dimension. (**cosygpmfphpp**)[9–12]

**gHSQC** experiments were used to acquire multiplicity edited HSQC spectra. A standard acquisition was as follows: si = 2048, ni = 256, ns = 8, ds = 32, rd = 1.0 s. The spectra were zero filled to 8k in the F2 and 1k in the F1 dimension. (**hsqcedetgpsisp2.3**)[13–18]

**gHMBC** experiments were used to acquire HSQC spectra. A standard acquisition was as follows: si = 2048, ni = 256, ns = 8, ds = 32, rd = 1.0 s, $J_{HMBC}$=8 Hz. The spectra were zero filled to 8k in the F2 and 1k in the F1 dimension. (**hmbcetgpl3nd**)[19,20]

**$S^3$ HMBC homo** experiments were used to acquire $S^3$ edited HMBC spectra for extraction of homonuclear coupling constants. A standard acquisition was as follows: si = 4096, ni = 256, ns = 8 or 16, ds = 32, rd = 1.0 s, $J_{HMBC}$=8 Hz. The spectra were zero filled to 16k in the F2 and 1k in the F1 dimension.

**$S^3$ HMBC hetero** experiments were used to acquire $S^3$ edited HMBC spectra for extraction of heteronuclear coupling constants. A standard acquisition was as follows: si = 4096, ni = 256, ns = 8 or 16, ds = 32, rd = 1.0 s, $J_{HMBC}$=8 Hz. The spectra were zero filled to 16k in the F2 and 1k in the F1 dimension.

The new pulse sequences of **$S^3$ HMBC homo/hetero** are not included, but are provided upon request, and the acquisition data are given in the thesis.

## A2. Other equipment

Analytical RP-UPLC-MS analyses were performed on a Waer Aquity RP-UPLC system with a diode array detector (DAD) coupled to a SQD mass spectrometer. The ionization method was ESI. The column was a Aquity UPLC BEH C18 column (1.7 µm, 2.1×50mm, 0.6 mL/min, 65 °C) Eluents were water and MeCN (both added 0.1 % $HCO_2H$), and a linear gradient of 5-100 % MeCN over aprox. 2.6 min was used.

Analytical RP-UPLC-HRMS data were recorded on a Maxis 3G UHR-QTOF-MS by Bruker Daltonics with an UltiMate 3000 Rapid Separation LC by system Thermo Scientific Dionex. The ionization method was ESI and UV spectra from the DAD were collected at wavelengths from 200-700 nm. The column used was a

Kinetex C18 (2.6 µm, 100x2.1 mm, 0.4 mL/min, 40 °C). Solvents were MeCN and water (both added 20 mM $HCO_2H$), and a linear gradient of 10-100 % MeCN in 10 minutes followed by 100 % MeCN for 3 minutes was used.

Optical rotation was measured on a Perkin-Elmer 341polarimeter, with a 100 mm cell of 1 mL, with a sodium lamp (589 nm, 20 °C).

# A3. Software

## A3.1. DataAnalysis

To analyse obtained HPLC-DAD-MS spectra, the program DataAnalysis by Bruker was used. An internal standard was used to gain higher mass accuracy. SmartFormula was used to generate and evaluate constituent formulas for a given mass.

## A3.2. Gaussian

The program Gaussian version 09 revision B.01 by Gaussian was used for DFT calculations including optimizations and NMR calculations.[21]

## A3.3. Maestro

The modelling suite Maestro version 10.2.010 (2015) by Schrödinger was used for force field calculations.[22] The programs MacroModel version 10.8 and Desmond version 4.2 were used.[23,24]

## A3.4. Matlab

The program Matlab version R2015a (2015) by The MathWorks Inc. was used for scripting and automating processes. The scripts were all made in house utilizing the general available scripts imbedded in the program.

The script RBNMR by Niels Nyberg was used to load NMR spectral data into Matlab.

## A3.5. MestReNova

The obtained 1D and 2D NMR spectra from my external stay were processed in MestReNova v. 6.2.1 by MestReLab Research S. L.

## A3.6. MSPIN

The software MSPIN version 1.3.3-79 (2013) by MestReLab Research S. L. was generally used for RDC (SVD) back-calculations.[25] The program may also be used for NOE and *J* coupling constant calculation but this was not generally used, though it was used to check the HLA calculations of the scripts. The NOE calculations were not used as the program uses a matrix approach.

## A3.7. PALES

The RDC software PALES was initially used and compared to MSPIN.[26] The latter program was used throughout and the two programs were found to give comparable results.

## A3.8. Topspin

The obtained 1D and 2D NMR spectra were generally processed in Topspin version 3.1 (2012) by Bruker BioSpin.

# A4. NMR Assignments

## A4.1. Natural products

### A4.1.1. Homomorphosin A

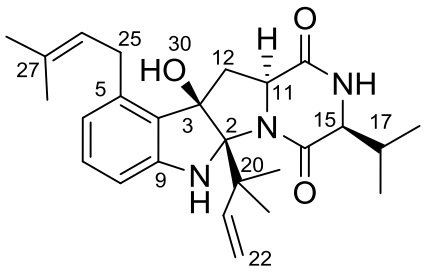



**Table A4-1. NMR assignment of homomorphosin A**

| # | $\delta_H$ [ppm] (int,mult,*J* [Hz]) | $\delta_C$ [ppm] |
|---|---|---|
| 1 | 6.72 (1H,s) | - |
| 2 | - | 90.8 |
| 3 | - | 89.1 |
| 4 | - | 127.3 |
| 5 | - | 139.3 |
| 6 | 6.47 (1H,d,7.6) | 119.3 |
| 7 | 6.99 (1H,t,7.7) | 129.4 |
| 8 | 6.56 (1H,d,7.8) | 107.8 |
| 9 | - | 149.6 |
| 10 | - | |
| 11 | 3.66 (1H,dd,11.5,7.3) | 57.0 |
| 12a | 2.54 (1H,m) | 34.6 |
| 12b | 2.62 (1H,dd,13.0,7.4) | 34.6 |
| 13 | - | 170.4 |
| 14 | 7.95 (1H,s) | - |
| 15 | 3.77 (1H,d,1.9) | 60.4 |
| 16 | - | 171.8 |
| 17a | 2.32 (1H,septd,7.1,2.0) | 26.7 |
| 17b | - | - |
| 18 | 0.86 (3H,d,6.8) | 15.8 |
| 19 | 0.98 (3H,d,7.2) | 18.1 |
| 20a | - | 44.6 |
| 20b | - | - |
| 21 | 6.27 (1H,dd,17.6,10.8) | 145.2 |
| 22a | 4.85 (1H,m) | 110.6 |
| 22b | 4.85 (1H,m) | 110.6 |
| 23 | 1.17 (3H,s) | 24.2 |
| 24 | 1.20 (3H,s) | 24.5 |
| 25a | 3.33 (1H,m) | 28.6 |

**Table A4-2. ROE intensities of homomorphosin A ($\tau_m$ = 150 ms)**

| H1 | H2 | $\eta^{norm}$ | Rel |
|---|---|---|---|
| 1 | 8 | 0.0135 | 0.74 |
| 1 | 21 | 0.0075 | 0.41 |
| 1 | 23/24 | 0.0185 | 1.01 |
| 1 | 23/24 | 0.0185 | 1.01 |
| 6 | 25a | 0.0092 | 0.50 |
| 6 | 26 | 0.0091 | 0.50 |
| 6 | 29 | 0.0017 | 0.09 |
| 7 | 8 | 0.0183 | 1.00 |
| 11 | 16 | 0.0162 | 0.89 |
| 15 | 17 | 0.0065 | 0.36 |
| 15 | 16 | 0.0105 | 0.57 |
| 16 | 17 | 0.0327 | 1.79 |
| 16 | 18 | 0.0079 | 0.43 |
| 16 | 19 | 0.0291 | 1.59 |
| 17 | 18 | 0.0259 | 1.42 |
| 17 | 19 | 0.0298 | 1.63 |
| 21 | 12a | 0.0076 | 0.42 |
| 21 | 23/24 | 0.0288 | 1.57 |
| 21 | 23/24 | 0.0288 | 1.57 |
| 21 | 30 | 0.0094 | 0.51 |
| 22b | 18 | -0.0012 | -0.06 |
| 22b | 23/24 | -0.0316 | -1.73 |
| 22b | 23/24 | -0.0316 | -1.73 |
| 25b | 12b | -0.0395 | -2.16 |
| 25b | 29 | -0.0140 | -0.77 |
| 26 | 28 | 0.0191 | 1.04 |
| 30 | 12a | 0.0081 | 0.45 |
| 30 | 23/24 | 0.0293 | 1.60 |
| 30 | 23/24 | 0.0293 | 1.60 |
| 30 | 25b | 0.0087 | 0.48 |

**Table A4-3. Experimental distances in Å from relative ROE intensities and *J*-coupling constants in Hz compared to best fit of multiple structures for the assignment used and by interchanging the prochiral methyl groups 18 and 19, or the diastereotopic protons of C25 ($\tau_m$ = 150 ms)**

| **NOE** | | | | **Right** | | **Dia-CH3-18/19** | | | **Dia-25** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| H1 | H2 | Exp | Calc | Diff | % | Calc | Diff | % | Calc | Diff | % |
| 15 | 17 | 2.94 | 2.93 | 0.01 | 0.5 | 2.59 | 0.36 | 12.2 | 2.91 | 0.03 | 1.1 |
| 15 | 16 | 2.72 | 2.73 | 0.00 | 0.1 | 2.72 | 0.00 | 0.1 | 2.73 | 0.00 | 0.1 |
| **7** | **8** | **2.48** | **2.48** | **0.00** | **0.0** | **2.48** | **0.00** | **0.0** | **2.48** | **0.00** | **0.0** |
| 1 | 21 | 2.88 | 2.96 | 0.08 | 2.7 | 2.86 | 0.02 | 0.6 | 3.03 | 0.15 | 5.2 |
| 1 | 23/24 | 3.34 | 3.37 | 0.03 | 0.9 | 3.36 | 0.02 | 0.7 | 3.36 | 0.02 | 0.7 |
| 6 | 25a | 2.78 | 2.76 | 0.02 | 0.9 | 2.74 | 0.04 | 1.4 | 2.72 | 0.06 | 2.3 |
| 6 | 26 | 2.79 | 2.80 | 0.01 | 0.3 | 2.84 | 0.05 | 1.7 | 2.76 | 0.03 | 1.2 |
| 6 | 29 | 4.41 | 4.42 | 0.01 | 0.1 | 4.42 | 0.01 | 0.1 | 4.44 | 0.03 | 0.6 |
| 30 | 23/24 | 3.09 | 3.35 | 0.26 | 8.5 | 3.37 | 0.28 | 8.9 | 3.35 | 0.26 | 8.4 |
| 30 | 12a | 2.84 | 2.75 | 0.09 | 3.1 | 2.76 | 0.08 | 2.7 | 2.64 | 0.20 | 6.9 |
| 30 | 25b | 2.81 | 2.92 | 0.11 | 4.0 | 2.91 | 0.10 | 3.7 | 2.89 | 0.08 | 2.8 |
| 26 | 28 | 2.96 | 2.82 | 0.14 | 4.6 | 2.82 | 0.14 | 4.7 | 2.82 | 0.14 | 4.6 |
| 22b | 23/24 | 3.05 | 3.16 | 0.11 | 3.6 | 3.16 | 0.11 | 3.6 | 3.15 | 0.10 | 3.4 |
| 21 | 23/24 | 3.10 | 3.28 | 0.18 | 5.8 | 3.28 | 0.18 | 5.7 | 3.30 | 0.20 | 6.3 |
| 21 | 12a | 2.87 | 2.96 | 0.09 | 3.1 | 2.92 | 0.05 | 1.7 | 3.02 | 0.15 | 5.2 |
| 21 | 30 | 2.77 | 2.89 | 0.12 | 4.2 | 2.88 | 0.11 | 3.9 | 2.92 | 0.15 | 5.4 |
| 22b | 18 | 4.71 | 4.67 | 0.04 | 0.8 | 4.69 | 0.02 | 0.4 | 4.63 | 0.08 | 1.6 |
| 16 | 18 | 3.42 | 3.50 | 0.08 | 2.3 | 3.21 | 0.22 | 6.3 | 3.49 | 0.07 | 2.0 |
| 16 | 19 | 2.76 | 3.01 | 0.25 | 9.1 | 3.18 | 0.42 | 15.3 | 3.01 | 0.26 | 9.3 |
| 16 | 17 | 2.25 | 2.52 | 0.27 | 12.1 | 2.52 | 0.27 | 12.1 | 2.52 | 0.27 | 12.1 |
| 25b | 29 | 3.11 | 3.07 | 0.04 | 1.3 | 3.22 | 0.10 | 3.3 | 3.40 | 0.29 | 9.2 |
| 25b | 12b | 2.18 | 2.27 | 0.09 | 4.1 | 2.33 | 0.15 | 6.8 | 2.21 | 0.03 | 1.2 |
| 11 | 16 | 2.53 | 2.44 | 0.09 | 3.7 | 2.44 | 0.10 | 3.8 | 2.44 | 0.10 | 3.8 |
| 17 | 18 | 2.81 | 2.62 | 0.20 | 7.0 | 2.61 | 0.20 | 7.2 | 2.62 | 0.20 | 7.0 |
| 17 | 19 | 2.75 | 2.61 | 0.14 | 5.0 | 2.61 | 0.14 | 5.0 | 2.61 | 0.14 | 5.0 |
| 1 | 8 | 2.61 | 2.66 | 0.05 | 1.9 | 2.66 | 0.05 | 1.8 | 2.66 | 0.05 | 1.9 |
| | | | MAE | 0.1 | 3.5 | | 0.1 | 4.4 | | 0.1 | 4.1 |

| ***J*-coupl** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| H1 | H2 | Exp | Calc | Diff | | Calc | Diff | | Calc | Diff |
| 12b | 11 | 7.3 | 6.5 | 0.8 | | 6.5 | 0.8 | | 6.6 | 0.8 |
| 15 | 17 | 2.0 | 1.9 | 0.1 | | 1.9 | 0.1 | | 1.9 | 0.1 |
| 12a | 11 | 11.5 | 10.2 | 1.3 | | 10.2 | 1.3 | | 10.2 | 1.3 |
| | | | MAE | 0.2 | | | 0.2 | | | 0.2 |

## A4.1.2. Cyclomorphosin A

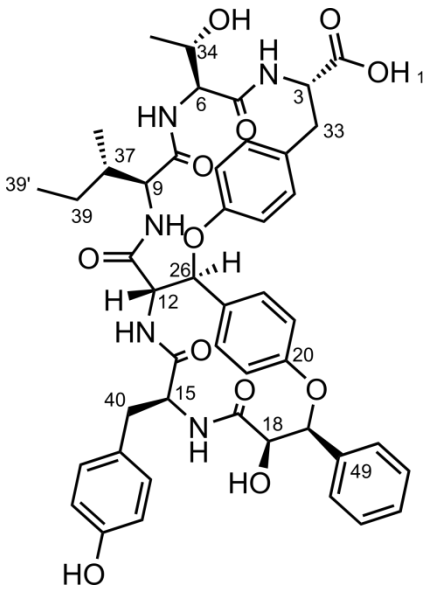
**Table A4-4. NMR assignment of cyclomorphosin A**

| # | $\delta_H$ [ppm] (int,mult,*J* [Hz]) | $\delta_C$ [ppm] | COSY | HMBC | NOESY (800 ms) |
|---|---|---|---|---|---|
| **1** | 12.76 (1H, s) | | | | 39,47,48 |
| **2** | | 172.6 | | | |
| **3** | 4.60 (1H, dt = 15.4,7.1) | 51 | 4,33a,33b | 2,33 | 33a,33b |
| **4** | 8.03 (1H, d = 8.7) | | 3 | 3,5 | 3,6,9,1,33a,33b,34a,35a,34b,35b |
| **5** | | 170.6 | | | |
| **6** | 4.75 (1H, d = 6.7) | 58.9 | 34a | 34,35,36 | 9, 37, 38,34a,35a,34b,35b,36a |
| **7** | - | | | | |
| **8** | | 165.8 | | | |
| **9** | 3.73 (1H, t=4.4) | 55.9 | 10,37 | 37,38 | 38,34a,35a,34b,35b |
| **10** | 7.60 (1H, d = 4.6) | | 9 | 8, 9,11 | 9,12,15,38 |
| **11** | | 168.2 | | | |
| **12** | 4.90 (1H, t = 9.6) | 58 | 13,26 | 11,14,26 | |
| **13** | 8.21 (1H, d = 10.2) | | 12 | 14 | 10, 12, 15, 22, 26,40a,40b |
| **14** | | 169.4 | | | |
| **15** | 4.14 (1H, q=7.9) | 52.5 | 16,40a,40b | 14,40 | 40a,40b |
| **16** | 6.56 (1H, d=9.1) | | 15 | 15,17 | 15, 18,40a,40b |
| **17** | | 169.9 | | | |
| **18** | 3.75 (1H, d=5.5) | 74.1 | 48 | 17,49 | |
| **19** | 5.95 (1H, s) | 78.2 | | 17,41,49 | 18,36a |
| **20** | | 155.7 | | | |
| **21** | 7.04 (1H, m) | 116.3 | 22 | 23,25 | 19 |
| **22** | 7.19 (1H, dd=8.1,2.6) | 129.6 | 21 | 20,24,26 | 15,19,26 |
| **23** | 7.02 (1H, m) | 120.3 | 24 | 21,25 | 18,19 |
| **24** | 7.42 (1H, m) | 127.6 | 23 | 20,26,31,49 | 12,29 |
| **25** | | 131.6 | | | |
| **26** | 5.44 (1H, d = 8.9) | 79.4 | 12 | 11,12,22,25 | |
| **27** | | 155 | | | |
| **28** | 6.82 (1H, dd=8.2,2.4) | 118.8 | 29 | 22,27,30 | 26 |
| **29** | 7.01a (1H, m) | 127.8 | 31,30 | 27,33 | 3, 26,33a,33b |
| **30** | 7.01b (1H, m) | 113.8 | 28 | 28 | 3, 26,33a,33b |
| **31** | 6.93 (1H, dd=8.3,1.1) | 129.5 | 29,30 | 27, 29,33 | 3,33a,33b |
| **32** | | 129.6 | | | |
| **33a** | 2.82 (1H, dd = 14.3,11.9) | 34.9 | 3,33b | 3,22,29 | |
| **33b** | 3.18 (1H, dd = 14.3,4.1) | 34.9 | 3,33a | 3,22 | 33a |
| **34a** | 1.92 (1H, m) | 20.8 | 6,34b,35b,36a,36b | 5,35,36 | |
| **35a** | 1.92 (1H, m) | 30.3 | 6,34b,35b,36a,36b | 5,34,36 | |
| **34b** | 1.79 (1H, m) | 20.8 | 34a,35a,36a | 5,6,36 | |
| **35b** | 1.79 (1H, m) | 30.3 | 34a,35a,36a | 5,6,36 | |
| **36a** | 3.38 (1H, m) | 45.9 | 34a,34b,36b | 34 | 38,34a,35a,34b,35b |
| **36b** | 3.52 (1H, m) | 45.9 | 34a,36a | 34,35 | 34a,35a,34b,35b |
| **37** | 4.19 (1H, m) | 64.8 | 9,38 | | 9,38 |
| **38** | 0.85 (3H, d= 6.4) | 17.4 | 37 | 9,37 | |
| **39** | 4.93 (1H, s) | | 37 | | 1,47,48 |
| **40a** | 2.31 (1H, td = 13.7,7.0) | 38.3 | 15,40b | 14,15,31,41 | |
| **40b** | 2.49 (1H, m) | 38.3 | 15,40a | 14,15,31,41 | |
| **41** | | 126.5 | | | |
| **42** | 6.78 (1H, d=8.4) | 129.5 | 43,45 | 22, 43, 46,40 | 15,40a,40b |
| **43** | 6.51 (1H, m) | 114.4 | | 41,43,46 | 38,36a |
| **44** | 6.78 (1H, d=8.4) | 129.5 | 43,45 | 22, 43, 46,40 | 15,40a,40b |
| **45** | 6.50 (1H, m) | 114.4 | 42,44 | 41,43,46 | 38,36a |
| **46** | | 155.5 | | | |
| **47** | 9.13 (1H, s) | | | 43 | 12,43,48 |
| **48** | 6.02 (1H, d = 4.5) | | 18 | 17,18 | |
| **49** | | 140.4 | | | |
| **50** | 7.59 (1H, m) | 126.5 | 53 | 19 | 18,19,53 |
| **51** | 7.41 (1H, m) | 127.7 | 54 | 20,26,31,49 | 12,29 |
| **52** | 7.59 (1H, m) | 126.5 | 53 | 19 | 18,19,53 |
| **53** | 7.39 (1H, m) | 127.7 | 50,52 | 20,26,31,49 | 12,29 |
| **54** | 7.33 (1H, t=7.2) | 127.3 | 51 | | |

## All distances

**Table A4-5. NOE intensities of cyclomorphosin A ($\tau_m$ = 200 ms)**

| H1 | H2 | $\eta^{norm}$ | Rel |
|---|---|---|---|
| 3 | 29 | 0.0403 | 0.11 |
| 3 | 31 | 0.0125 | 0.03 |
| 3 | 33b | 0.0401 | 0.11 |
| 4 | 6 | 0.0497 | 0.14 |
| 4 | 33a | 0.0153 | 0.04 |
| 4 | 33b | 0.0029 | 0.01 |
| 4 | 35b | 0.0061 | 0.02 |
| 4 | 36b | 0.0081 | 0.02 |
| 6 | 9 | 0.1048 | 0.29 |
| 6 | 34a | 0.0547 | 0.15 |
| 6 | 34b | 0.0343 | 0.09 |
| 9 | 10 | 0.0206 | 0.06 |
| 9 | 37 | 0.0306 | 0.08 |
| 10 | 12 | 0.0679 | 0.19 |
| 10 | 13 | 0.0121 | 0.03 |
| 10 | 38 | 0.0115 | 0.03 |
| 12 | 24 | 0.0361 | 0.10 |
| 13 | 15 | 0.0964 | 0.27 |
| 13 | 22 | 0.0117 | 0.03 |
| 13 | 26 | 0.0237 | 0.07 |
| 15 | 16 | 0.0110 | 0.03 |
| 15 | 40a | 0.0288 | 0.08 |
| 15 | 40b | 0.0212 | 0.06 |
| 16 | 18 | 0.0085 | 0.02 |
| 16 | 40a | 0.0131 | 0.04 |
| 18 | 19 | 0.0411 | 0.11 |
| 19 | 21 | 0.1532 | 0.42 |
| 22 | 26 | 0.0696 | 0.19 |
| 26 | 30 | 0.1017 | 0.28 |
| 29 | 33a | 0.0241 | 0.07 |
| 29 | 33b | 0.0297 | 0.08 |
| 31 | 33a | 0.0599 | 0.16 |
| 33a | 33b | 0.2816 | 0.77 |
| 34b | 36b | 0.0537 | 0.15 |
| 35b | 36a | 0.0689 | 0.19 |
| 37 | 38 | 0.0288 | 0.08 |
| 38 | 43/45 | 0.0047 | 0.01 |
| 40a | 40b | 0.3635 | 1.00 |



## Ring A distances

**Table A4-6. Experimental distances in Å from relative NOE intensities compared to best fit of multiple structures for possible diastereomers of ring A, cyclomorphosin A ($\tau_m$ = 200 ms)**

| Exp. | 12S,15S,18R, 19S,26S | 12S,15R,18R, 19S,26S | 12S,15S,18S, 19S,26S | 12S,15R,18S, 19S,26S | 12R,15S,18R, 19S,26R |
|---|---|---|---|---|---|
| 2.57 | 2.76 | 2.71 | 2.82 | 2.65 | 2.89 |
| 2.18 | 2.12 | 2.45 | 2.12 | 2.48 | 2.39 |
| 3.10 | 3.20 | 3.44 | 3.21 | 3.27 | 3.22 |
| 2.75 | 2.91 | 2.85 | 2.93 | 2.97 | 3.02 |
| 3.13 | 2.99 | 2.90 | 2.94 | 2.93 | 2.97 |
| 2.67 | 2.69 | 2.72 | 2.70 | 2.74 | 2.62 |
| 2.80 | 2.80 | 2.73 | 2.79 | 2.73 | 2.78 |
| 3.27 | 3.31 | 2.66 | 2.16 | 2.79 | 3.35 |
| 3.04 | 2.92 | 2.99 | 2.89 | 3.03 | 2.67 |
| 2.51 | 2.65 | 2.55 | 3.09 | 2.65 | 2.68 |
| 2.02 | 2.14 | 2.29 | 2.15 | 2.45 | 2.49 |
| 2.30 | 2.32 | 2.32 | 2.34 | 2.30 | 2.58 |
| **1.75** | **1.75** | **1.75** | **1.75** | **1.75** | **1.75** |

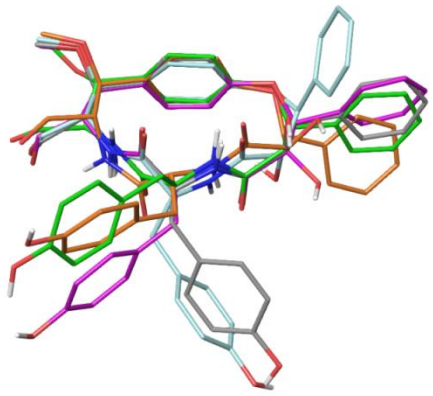


**Figure A4.1. Representative structures of ring A. Grey: 12*S*,15*S*,18*R*,19*S*,26*S*, orange: 12*S*,15*R*,18*R*,19*S*,26*S*, turquoise: 12*S*,15*R*,18*S*,19*S*,26*S*, purple: 12*S*,15*S*,18*S*,19*S*,26*S*, green: 12*R*,15*S*,18*R*,19*S*,26*R*,**

## Ring B distances

**Table A4-7. Experimental distances in Å from relative NOE intensities compared to best fit of multiple structures for possible diastereomers of ring B, cyclomorphosin A ($\tau_m$ = 200 ms)**

| Exp. | 3S,6S,9S, 12S,26S | 3R,6S,9S, 12S,26S | 3S,6R,9S, 12S,26S | 3S,6S,9R, 12S,26S | 3R,6R,9S, 12S,26S | 3R,6S,9R, 12S,26S | 3S,6R,9R, 12S,26S | 3R,6R,9R, 12S,26S |
|---|---|---|---|---|---|---|---|---|
| 2.42 | 2.58 | 2.82 | 2.53 | 2.78 | 3.02 | 2.73 | 2.56 | 2.70 |
| 2.94 | 2.96 | 3.02 | 2.93 | 3.05 | 3.24 | 2.95 | 2.92 | 3.01 |
| 2.42 | 2.56 | 2.51 | 2.59 | 2.50 | 2.50 | 2.54 | 2.51 | 2.55 |
| 2.33 | 2.37 | 2.52 | 2.71 | 2.23 | 2.38 | 2.63 | 2.48 | 2.27 |
| 2.84 | 2.77 | 2.62 | 2.65 | 2.65 | 2.63 | 2.71 | 2.72 | 2.67 |
| 3.75 | 3.83 | 3.75 | 3.76 | 3.72 | 3.69 | 3.80 | 3.80 | 3.78 |
| 2.06 | 2.26 | 2.24 | 2.39 | 2.50 | 2.53 | 2.32 | 2.43 | 2.39 |
| 2.70 | 2.81 | 2.77 | 2.78 | 2.59 | 2.67 | 2.64 | 2.60 | 2.60 |
| 2.53 | 2.53 | 2.57 | 2.54 | 2.57 | 2.51 | 2.62 | 2.68 | 2.65 |
| 2.22 | 2.24 | 2.21 | 2.29 | 2.20 | 2.34 | 2.16 | 2.32 | 2.31 |
| 2.96 | 2.94 | 2.89 | 2.49 | 2.82 | 2.91 | 3.05 | 2.94 | 2.87 |
| 2.98 | 3.27 | 3.19 | 3.32 | 3.33 | 3.40 | 3.14 | 3.20 | 3.28 |
| 2.07 | 2.20 | 2.32 | 2.23 | 2.18 | 2.46 | 2.49 | 2.29 | 2.44 |
| 2.63 | 2.86 | 2.69 | 2.91 | 2.88 | 2.73 | 2.77 | 2.91 | 2.72 |
| 2.54 | 2.84 | 3.08 | 2.79 | 2.75 | 2.81 | 2.85 | 2.84 | 2.87 |
| 2.26 | 2.48 | 2.68 | 2.49 | 2.49 | 2.60 | 2.59 | 2.50 | 2.55 |
| **1.75** | **1.75** | **1.75** | **1.75** | **1.75** | **1.75** | **1.75** | **1.75** | **1.75** |
| 2.56 | 2.62 | 2.61 | 2.61 | 2.62 | 2.62 | 2.62 | 2.61 | 2.62 |

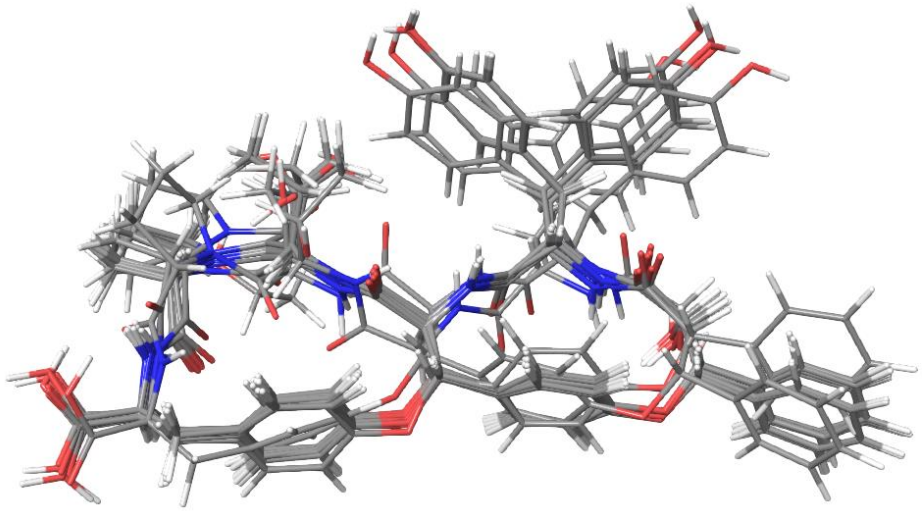


**Figure A4.2. Representative structures of the conformational space covered by CM A.**

## Only best structure used

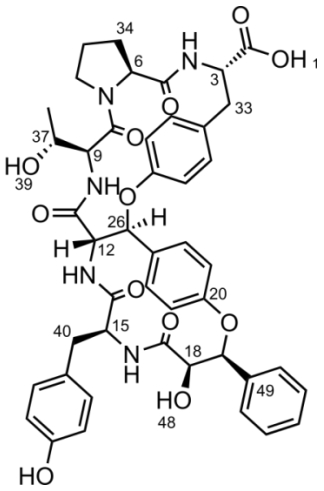
**Table A4-8. Lowest MAE found between experimental distances from relative NOE intensities and distances of the single structure for possible diastereomers of ring A and B, cyclomorphosin A ($\tau_m$ = 200 ms)**

| MAE | NOE [Å] | *J*s [Hz] |
|---|---|---|
| Ring A | | |
| 12S,15S,18R,19S,26S | 0.16 | 4.3 |
| 12S,15R,18R,19S,26S | 0.44 | 4.0 |
| 12S,15S,18S,19S,26S | 0.38 | 8.0 |
| 12S,15R,18S,19S,26S | 0.54 | 5.8 |
| 12R,15S,18R,19S,26R | 0.80 | 3.6 |
| Ring B | | |
| 3S,6S,9S,12S,26S | 0.70 | 0.4 |
| 3R,6S,9S,12S,26S | 0.84 | 1.2 |
| 3S,6R,9S,12S,26S | 0.76 | 1.0 |
| 3S,6S,9R,12S,26S | 0.53 | 1.6 |
| 3R,6R,9S,12S,26S | 0.81 | 0.6 |
| 3R,6S,9R,12S,26S | 0.69 | 2.2 |
| 3S,6R,9R,12S,26S | 0.57 | 1.7 |
| 3R,6R,9R,12S,26S | 0.65 | 1.4 |

## A4.1.3. Cyclomorphosin B

**Table A4-9. NMR assignment of cyclomorphosin B**

| # | $\delta_H$ [ppm] (int,mult,*J* [Hz]) | $\delta_C$ [ppm] | COSY | HMBC | NOESY (800 ms) |
|---|---|---|---|---|---|
| **1** | - | | | | |
| **2** | | | | | |
| **3** | 4.50 (1H,td=10.1,3.6) | 52.2 | 4,33a | | 33b |
| **4** | 7.83 (1H,d=9.4) | | 3 | | 6 |
| **5** | | 168.3 | | | |
| **6** | 4.19 (1H,dd=8.0,6.2) | 57.5 | 7 | 5, 8,34,35 | 4,7 |
| **7** | 6.87 (1H,d=6.8) | | | | 10 |
| **8** | | 170 | | | |
| **9** | 4.13 (1H,t=8.7) | 58.1 | 10,37 | 8, 37,39 | 10,38 |
| **10** | 8.53 (1H,d=9.4) | | 9 | 11 | 7,12 |
| **11** | | 169.2 | | | |
| **12** | 4.66 (1H,dd=9.7,7.3) | 58.6 | 13,26 | 14,26 | 10,22 |
| **13** | 8.33 (1H,d=9.9) | | 12 | 14 | 15,21 |
| **14** | | 169.3 | | | |
| **15** | 4.39 (1H,dt=9.3,8.0) | 51.7 | 16,40a,40b | | 13, 16,40b |
| **16** | 6.48 (1H,d=9.5) | | 15 | | 15,48 |
| **17** | | 169.7 | | | |
| **18** | 3.75b (1H,m) | 74.2 | 48 | | |
| **19** | 5.93 (1H,s) | 78.7 | | 49,50 | 21,50 |
| **20** | | 155.6 | | | |
| **21** | 7.06 (1H,d=8.5) | 120.6 | 22 | 23,25 | 13,19,22,26 |
| **22** | 7.46 (1H,d=8.5) | 126.8 | 21 | 24 | 12,21 |
| **23** | 7.04 (1H,m) | 116.9 | | 20,25 | |
| **24** | 7.01 (1H,m) | 128.9 | 23 | 22 | |
| **25** | | 132.2 | | | |
| **26** | 5.53 (1H,d=7.2) | 78.9 | 12 | 11,22,24,25,27 | 21,28 |
| **27** | | 155.2 | | | |
| **28** | 6.76 (1H,d=6.8) | | 29 | | 26 |
| **29** | 6.98 (1H,m) | | 28 | | |
| **30** | 6.76 (1H,d=6.8) | | 29 | | |
| **31** | 6.98 (1H,m) | | 28 | | |
| **32** | | | | | |
| **33a** | 2.54 (1H,m) | 36.4 | 3,33b | | 33b |
| **33b** | 3.09 (1H,dd=13.3,3.6) | 36.4 | 33a | | 3,33a |
| **34** | 3.75a (1H,m) | 65.8 | 35,36 | | 6,16,19,35,48 |
| **35** | 0.95 (3H,d=6.2) | 19.7 | | 6,34 | |
| **36** | 4.61 (1H,d=5.1) | | | 34,35 | |
| **37** | 1.66 (1H,m) | 36.5 | 9,38 | | |
| **38** | 0.75 (1H,d=6.7) | 15.1 | 37 | 9, 37,39 | |
| **39a** | 0.98 (1H,m) | 24.4 | 39b,39’ | | |
| **39b** | 1.30 (1H,m) | 24.4 | 39a,39’ | | |
| **39’** | 0.74 (1H,d=7.3) | 15.1 | 39a,39b | 37,39 | |
| **40a** | 2.33 (1H,d=13.9,7.2) | 37.9 | 15,40b | 14,15,41,42 | |
| **40b** | 2.49 (1H,m) | 37.9 | 15,40a | 41,42 | |
| **41** | | 126.5 | | | |
| **42** | 6.80 (1H,d=8.4) | 129.6 | 43 | 42,46 | 43 |
| **43** | 6.51 (1H,d=8.3) | 114.4 | 42 | 41,43,46 | 42 |
| **44** | 6.80 (1H,d=8.4) | 129.6 | 43 | 42,46 | 43 |
| **45** | 6.51 (1H,d=8.3) | 114.4 | 42 | 41,43,46 | 42 |
| **46** | | 155.4 | | | |
| **47** | 9.08 (1H,s) | | | 43,46 | |
| **48** | 6.01 (1H,d=6.6) | | | 17,18,19 | 16 |
| **49** | | 140.6 | | | |
| **50** | 7.59 (1H,d=7.6) | 126.6 | 51 | 19,50,54 | 51 |
| **51** | 7.40 (1H,t=7.5) | 127.9 | 50,54 | 49,51 | 50,54 |
| **52** | 7.59 (1H,d=7.6) | 126.6 | 51 | 19,50,54 | 51 |
| **53** | 7.40 (1H,t=7.5) | 127.9 | 50,54 | 49,51 | 50,54 |
| **54** | 7.32 (1H,t=7.3) | 127.2 | 51 | 50 | 51 |



**Cyclomorphosin A and B, other data**

Cyclomorphosin A: HRMS: *m/z* = 850.3307. $[M+H]^+$, calculated for $[C_{46}H_{43}N_9O_8+H]^+$: *m/z* = 850.3307. Adducts: 872 $[M+Na]^+$, 425 $[M+2H]^{2+}$. $[\alpha]_{589.3}^{20}$ = -0.02°. $UV^{max}$: 225 nm, 276 nm.

Cyclomorphosin AB: HRMS: *m/z* = 866.36a1. $[M+H]^+$, calculated for $[C_{47}H_{47}N_9O_8+H]^+$: *m/z* = 866.36b0. Adducts: 888 $[M+Na]^+$, 883 $[M+NH_3]^+$, 433 $[M+2H]^{2+}$. $[\alpha]_{589.3}^{20}$ = 0.3°. $UV^{max}$: 226 nm, 271 nm.

### A4.1.4. Homomorphosterol

**Table A4-10. NMR assignment of homomorphosterol**

| # | $\delta_H$ [ppm] (int,mult,*J* [Hz]) | $\delta_C$ [ppm] |
|---|---|---|
| 1a | 1.06 (1H,dd, 13.3,2.7) | 35.5 |
| 1b | 1.66 (1H,m) | 35.5 |
| 2a | 1.32 (1H,m) | 31.4 |
| 2b | 1.72 (1H,m) | 31.4 |
| 3 | 3.25 (1H,m) | 68.9 |
| 4a | 2.08b (1H,m) | 41.9 |
| 4b | 2.29 (1H,dd, 13.9, 2.8) | 41.9 |
| 5 | | 143.5 |
| 6 | 5.29 (1H, br. s) | 120.4 |
| 7 | 2.58 (1H, br. s) | 35.5 |
| 8 | | 127.4 |
| 9 | 1.80 (1H,m) | 46.8 |
| 10 | | 36.3 |
| 11a | 1.48 (1H,m) | 19 |
| 11b | 1.52 (1H,m) | 19 |
| 12a | 1.30 (1H,m) | 38.3 |
| 12b | 1.97 (1H,12.4, 2.8) | 38.2 |
| 13 | | 42 |
| 14 | | 145.7 |
| 15 | 2.38 (1H, td, 10.1, 3.8) | 36 |
| 16a | 1.58 (1H,m) | 35.3 |
| 16b | 1.74 (1H,m) | 35.3 |
| 17 | 1.25 (1H,m) | 54.1 |
| 18 | 0.92 (3H,m) | 19.7 |
| 19 | 0.78 (3H,m) | 17.8 |
| 20 | 2.08a (1H,m) | 38.4 |
| 21 | 1.00 (3H,d, 6.6) | 20.6 |
| 22 | 5.19 (1H,m) | 135.1 |
| 23 | 5.25 (1H,m) | 131.6 |
| 24 | 1.87 (1H,q, 6.0) | 42.1 |
| 25 | 1.46 (1H,m) | 32.5 |
| 26 | 0.80b (3H,m) | 19.6 |
| 27 | 0.80a (3H,m) | 19.6 |
| 28 | 0.89 (3H,d, 7.0) | 17.2 |
| 30 | 2.16 (1H, dd, 5.7,10.1) | 49.9 |
| 31 | 2.91 (1H,dd, 6.9,5.7) | 47.2 |
| 32 | | 175 |
| 33 | | 176.6 |
| OH | 4.64 (1H,s) | - |
| COOH | 12.12 (1H, br. s) | - |

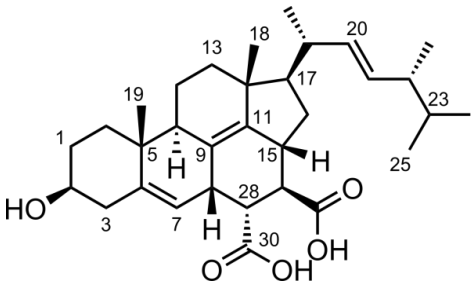


**Table A4-11. Chemical shifts of homomorphosterol compared to theoretical shielding constants of possible diastereomers, calculated by mPW1PW91/6-311+G(2d,p) using GIAO.**

| | **Shielding constants** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Exp. $^{1}$H δ (ppm) | (RRSS) | (RRRS) | (RSRS) | (RSSS) | (SRSR) | (SRRR) | (SSSR) | (SSRR) |
| 1.25 | -30.2 | -29.9 | -30.1 | -30.1 | -29.9 | -30.0 | -29.9 | -30.0 |
| 1.58 | -30.0 | -29.8 | -30.3 | -30.1 | -29.9 | -29.8 | -29.9 | -29.9 |
| 1.74 | -30.1 | -29.7 | -29.7 | -30.1 | -30.2 | -30.4 | -30.2 | -30.0 |
| 2.38 | -28.3 | -29.1 | -28.7 | -28.1 | -29.0 | -28.8 | -29.1 | -28.8 |
| 2.58 | -28.4 | -28.6 | -28.9 | -28.8 | -28.3 | -28.3 | -28.3 | -27.9 |
| 5.29 | -25.6 | -25.6 | -25.9 | -25.9 | -25.7 | -25.9 | -25.9 | -25.6 |
| 2.29 | -29.2 | -29.2 | -29.3 | -29.2 | -29.3 | -29.3 | -29.3 | -29.4 |
| 2.08 | -29.5 | -29.3 | -29.5 | -29.5 | -29.4 | -29.4 | -29.5 | -29.4 |
| 3.25 | -27.9 | -28.0 | -27.9 | -27.9 | -28.0 | -28.0 | -27.9 | -28.0 |
| 1.32 | -29.9 | -30.0 | -30.0 | -29.9 | -30.0 | -30.0 | -29.9 | -29.9 |
| 1.72 | -30.1 | -30.1 | -30.1 | -30.2 | -30.2 | -30.2 | -30.1 | -30.1 |
| 1.06 | -30.5 | -30.5 | -30.4 | -30.4 | -30.5 | -30.5 | -30.5 | -30.5 |
| 1.66 | -29.9 | -29.9 | -29.8 | -29.9 | -29.9 | -29.9 | -29.8 | -29.9 |
| 1.80 | -29.9 | -29.7 | -29.7 | -29.7 | -29.4 | -29.4 | -29.4 | -29.3 |
| 1.48 | -30.0 | -30.1 | -30.0 | -30.0 | -30.0 | -30.1 | -30.0 | -29.9 |
| 1.52 | -30.0 | -29.9 | -29.9 | -30.1 | -29.9 | -29.9 | -29.9 | -29.8 |
| 1.03 | -30.5 | -30.4 | -30.4 | -30.4 | -30.5 | -30.6 | -30.5 | -30.5 |
| 1.97 | -29.9 | -29.7 | -29.7 | -29.8 | -29.8 | -29.8 | -29.7 | -29.9 |
| 0.92 | -31.0 | -31.1 | -31.1 | -31.1 | -31.1 | -31.1 | -31.1 | -30.9 |
| 0.78 | -31.0 | -30.9 | -30.9 | -31.0 | -30.9 | -30.9 | -31.0 | -31.1 |
| 2.91 | -28.2 | -28.1 | -28.7 | -29.1 | -29.1 | -28.8 | -28.3 | -28.1 |
| 2.16 | -28.3 | -29.1 | -29.0 | -28.2 | -28.6 | -28.3 | -28.9 | -28.5 |
| a | 1.23 | 1.19 | 1.12 | 1.14 | 1.14 | 1.14 | 1.15 | 1.21 |
| b | -31.86 | -31.81 | -31.73 | -31.73 | -31.77 | -31.75 | -31.75 | -31.79 |

| | **Shielding constants** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Exp. $^{13}$C δ (ppm) | (RRSS) | (RRRS) | (RSRS) | (RSSS) | (SRSR) | (SRRR) | (SSSR) | (SSRR) |
| 35.3 | -146.6 | -147.0 | -145.4 | -148.1 | -148.4 | -149.7 | -148.3 | -149.2 |
| 36.0 | -147.4 | -141.7 | -145.4 | -147.0 | -138.5 | -143.3 | -141.1 | -139.1 |
| 145.7 | -32.1 | -30.3 | -29.1 | -34.1 | -31.5 | -34.3 | -31.5 | -36.2 |
| 127.4 | -54.5 | -48.2 | -48.7 | -49.6 | -48.6 | -48.2 | -49.9 | -51.0 |
| 35.5 | -142.4 | -144.5 | -142.7 | -142.8 | -147.2 | -147.5 | -145.7 | -145.7 |
| 120.4 | -56.2 | -59.1 | -59.3 | -59.2 | -56.4 | -57.5 | -59.1 | -56.3 |
| 143.5 | -31.3 | -30.0 | -29.1 | -29.6 | -30.5 | -30.1 | -29.0 | -34.4 |
| 41.9 | -144.7 | -144.2 | -144.6 | -144.1 | -144.3 | -144.4 | -144.4 | -144.7 |
| 68.9 | -112.4 | -112.1 | -112.6 | -112.0 | -111.8 | -111.5 | -112.0 | -111.1 |
| 31.4 | -152.1 | -151.4 | -152.0 | -151.3 | -151.9 | -151.3 | -152.0 | -152.0 |
| 35.5 | -148.4 | -148.0 | -148.0 | -148.6 | -148.3 | -148.6 | -148.0 | -147.7 |
| 36.3 | -145.3 | -144.4 | -144.8 | -145.1 | -144.0 | -143.8 | -144.7 | -141.5 |
| 46.8 | -134.1 | -134.6 | -136.0 | -134.5 | -137.1 | -136.4 | -136.8 | -137.3 |
| 19.0 | -163.6 | -164.7 | -165.1 | -163.9 | -164.4 | -164.7 | -165.2 | -163.3 |
| 38.3 | -147.7 | -147.6 | -148.2 | -147.4 | -149.3 | -148.8 | -148.6 | -149.8 |
| 42.0 | -138.8 | -139.4 | -139.6 | -139.0 | -142.1 | -139.8 | -141.2 | -140.7 |
| 19.7 | -166.0 | -167.7 | -168.0 | -166.3 | -167.0 | -167.1 | -166.6 | -167.9 |
| 17.8 | -170.2 | -168.7 | -169.8 | -170.1 | -167.5 | -167.5 | -167.1 | -167.6 |
| 47.2 | -139.7 | -135.1 | -134.8 | -138.7 | -133.6 | -132.9 | -132.1 | -134.4 |
| 49.9 | -136.2 | -134.0 | -133.1 | -134.1 | -130.8 | -135.8 | -131.5 | -141.4 |
| a | 1.06 | 1.07 | 1.08 | 1.06 | 1.07 | 1.07 | 1.07 | 1.05 |
| b | -185.79 | -185.66 | -186.26 | -185.79 | -185.68 | -185.96 | -185.53 | -185.05 |

**Table A4-12. Chemical shifts of homomorphosterol compared to theoretical chemical shifts, determined by linear scaling of the shielding tensors, of possible diastereomers, calculated by mPW1PW91/6-311+G(2d,p) using GIAO. Differences given in parentheses.**

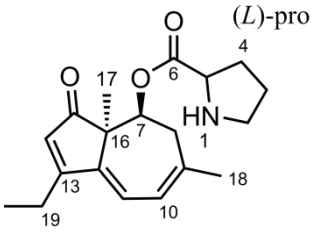
| | Theoretical $^{1}$H δ (ppm) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Exp. $^{1}$H δ (ppm) | (RRSS) | (RRRS) | (RSRS) | (RSSS) | (SRSR) | (SRRR) | (SSSR) | (SSRR) |
| 1.25 | 1.39 (0.14) | 1.64 (0.39) | 1.47 (0.22) | 1.46 (0.21) | 1.61 (0.36) | 1.56 (0.31) | 1.66 (0.41) | 1.49 (0.24) |
| 1.58 | 1.54 (0.04) | 1.7 (0.12) | 1.31 (0.27) | 1.43 (0.15) | 1.63 (0.05) | 1.7 (0.12) | 1.66 (0.08) | 1.59 (0.01) |
| 1.74 | 1.48 (0.26) | 1.77 (0.03) | 1.81 (0.07) | 1.46 (0.28) | 1.34 (0.4) | 1.22 (0.52) | 1.32 (0.42) | 1.46 (0.28) |
| 2.38 | 2.88 (0.5) | 2.31 (0.07) | 2.72 (0.34) | 3.2 (0.82) | 2.42 (0.04) | 2.56 (0.18) | 2.34 (0.04) | 2.5 (0.12) |
| 2.58 | 2.8 (0.22) | 2.74 (0.16) | 2.56 (0.02) | 2.59 (0.01) | 3.04 (0.46) | 3 (0.42) | 2.98 (0.4) | 3.2 (0.62) |
| 5.29 | 5.07 (0.22) | 5.23 (0.06) | 5.2 (0.09) | 5.15 (0.14) | 5.3 (0.01) | 5.16 (0.13) | 5.1 (0.19) | 5.09 (0.2) |
| 2.29 | 2.19 (0.1) | 2.18 (0.11) | 2.21 (0.08) | 2.27 (0.02) | 2.14 (0.15) | 2.14 (0.15) | 2.14 (0.15) | 1.95 (0.34) |
| 2.08 | 1.9 (0.18) | 2.1 (0.02) | 1.98 (0.1) | 2.01 (0.07) | 2.03 (0.05) | 2.04 (0.04) | 2 (0.08) | 1.94 (0.14) |
| 3.25 | 3.26 (0.01) | 3.21 (0.04) | 3.41 (0.16) | 3.34 (0.09) | 3.31 (0.06) | 3.3 (0.05) | 3.35 (0.1) | 3.16 (0.09) |
| 1.32 | 1.6 (0.28) | 1.51 (0.19) | 1.59 (0.27) | 1.6 (0.28) | 1.57 (0.25) | 1.53 (0.21) | 1.6 (0.28) | 1.54 (0.22) |
| 1.72 | 1.41 (0.31) | 1.41 (0.31) | 1.42 (0.3) | 1.36 (0.36) | 1.35 (0.37) | 1.37 (0.35) | 1.41 (0.31) | 1.37 (0.35) |
| 1.06 | 1.13 (0.07) | 1.09 (0.03) | 1.17 (0.11) | 1.2 (0.14) | 1.14 (0.08) | 1.09 (0.03) | 1.09 (0.03) | 1.1 (0.04) |
| 1.66 | 1.62 (0.04) | 1.58 (0.08) | 1.69 (0.03) | 1.61 (0.05) | 1.66 (0) | 1.62 (0.04) | 1.72 (0.06) | 1.58 (0.08) |
| 1.8 | 1.63 (0.17) | 1.8 (0) | 1.84 (0.04) | 1.81 (0.01) | 2.03 (0.23) | 2.05 (0.25) | 2.04 (0.24) | 2.09 (0.29) |
| 1.48 | 1.52 (0.04) | 1.46 (0.02) | 1.59 (0.11) | 1.49 (0.01) | 1.55 (0.07) | 1.47 (0.01) | 1.54 (0.06) | 1.6 (0.12) |
| 1.52 | 1.49 (0.03) | 1.57 (0.05) | 1.66 (0.14) | 1.44 (0.08) | 1.62 (0.1) | 1.59 (0.07) | 1.58 (0.06) | 1.61 (0.09) |
| 1.3 | 1.13 (0.17) | 1.2 (0.1) | 1.18 (0.12) | 1.22 (0.08) | 1.11 (0.19) | 1.06 (0.24) | 1.12 (0.18) | 1.11 (0.19) |
| 1.97 | 1.59 (0.38) | 1.79 (0.18) | 1.78 (0.19) | 1.71 (0.26) | 1.69 (0.28) | 1.68 (0.29) | 1.76 (0.21) | 1.54 (0.43) |
| 0.92 | 0.72 (0.2) | 0.64 (0.28) | 0.57 (0.35) | 0.55 (0.37) | 0.63 (0.29) | 0.61 (0.31) | 0.54 (0.38) | 0.73 (0.19) |
| 0.78 | 0.75 (0.03) | 0.73 (0.05) | 0.72 (0.06) | 0.67 (0.11) | 0.74 (0.04) | 0.73 (0.05) | 0.65 (0.13) | 0.59 (0.19) |
| 2.91 | 3.02 (0.11) | 3.1 (0.19) | 2.73 (0.18) | 2.32 (0.59) | 2.38 (0.53) | 2.57 (0.34) | 2.99 (0.08) | 3.04 (0.13) |
| 2.16 | 2.92 (0.76) | 2.28 (0.12) | 2.43 (0.27) | 3.15 (0.99) | 2.76 (0.6) | 2.99 (0.83) | 2.46 (0.3) | 2.75 (0.59) |
| MAE | 0.19 | 0.12 | 0.16 | 0.23 | 0.21 | 0.22 | 0.19 | 0.22 |

| | Theoretical $^{13}$C δ (ppm) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Exp. $^{13}$C δ (ppm) | (RRSS) | (RRRS) | (RSRS) | (RSSS) | (SRSR) | (SRRR) | (SSSR) | (SSRR) |
| 35.3 | 36.9 (1.6) | 36 (0.7) | 37.7 (2.4) | 35.4 (0.1) | 34.6 (0.7) | 33.9 (1.4) | 34.7 (0.6) | 34.2 (1.1) |
| 36 | 36.2 (0.1) | 41 (5) | 37.8 (1.8) | 36.5 (0.4) | 44 (8) | 39.9 (3.9) | 41.6 (5.6) | 44 (8) |
| 145.7 | 146.9 (1.2) | 146 (0.3) | 147.6 (1.9) | 144 (1.7) | 146.2 (0.5) | 143.4 (2.3) | 146.2 (0.5) | 143.5 (2.2) |
| 127.4 | 125.3 (2.1) | 129.1 (1.7) | 129.1 (1.7) | 129.2 (1.8) | 129.9 (2.5) | 130.2 (2.8) | 128.6 (1.2) | 129.2 (1.8) |
| 35.5 | 40.9 (5.4) | 38.3 (2.8) | 40.3 (4.8) | 40.5 (4.9) | 35.8 (0.3) | 36 (0.4) | 37.2 (1.7) | 37.6 (2.1) |
| 120.4 | 123.7 (3.3) | 118.8 (1.6) | 119.1 (1.3) | 120.1 (0.4) | 122.4 (2) | 121.4 (1) | 119.8 (0.6) | 124.1 (3.7) |
| 143.5 | 147.6 (4.1) | 146.3 (2.8) | 147.6 (4.1) | 148.3 (4.8) | 147.2 (3.7) | 147.4 (3.9) | 148.6 (5.1) | 145.3 (1.7) |
| 41.9 | 38.7 (3.2) | 38.6 (3.3) | 38.5 (3.4) | 39.1 (2.8) | 38.5 (3.4) | 38.8 (3.1) | 38.5 (3.4) | 38.6 (3.3) |
| 68.9 | 69.8 (0.9) | 68.9 (0) | 68.7 (0.2) | 69.8 (0.9) | 69.5 (0.6) | 70.1 (1.1) | 69.3 (0.4) | 71.1 (2.2) |
| 31.4 | 31.6 (0.2) | 31.8 (0.4) | 31.5 (0.1) | 32.3 (0.9) | 31.3 (0.1) | 32.3 (0.9) | 31.2 (0.2) | 31.5 (0.1) |
| 35.5 | 35.2 (0.3) | 35 (0.5) | 35.3 (0.2) | 34.8 (0.7) | 34.6 (0.9) | 34.9 (0.6) | 35 (0.5) | 35.7 (0.2) |
| 36.3 | 38.1 (1.8) | 38.5 (2.2) | 38.3 (2) | 38.2 (1.9) | 38.8 (2.5) | 39.4 (3.1) | 38.1 (1.8) | 41.7 (5.4) |
| 46.8 | 48.9 (2.1) | 47.7 (0.9) | 46.6 (0.2) | 48.3 (1.5) | 45.4 (1.4) | 46.4 (0.4) | 45.7 (1.1) | 45.7 (1.1) |
| 19 | 20.6 (1.6) | 19.3 (0.3) | 19.2 (0.2) | 20.3 (1.3) | 19.3 (0.3) | 19.6 (0.6) | 18.6 (0.4) | 20.6 (1.6) |
| 38.3 | 35.8 (2.5) | 35.5 (2.8) | 35.1 (3.2) | 36 (2.3) | 33.7 (4.6) | 34.7 (3.6) | 34.4 (3.9) | 33.6 (4.7) |
| 42 | 44.4 (2.4) | 43.1 (1.1) | 43.3 (1.3) | 44 (2) | 40.6 (1.4) | 43.2 (1.2) | 41.5 (0.5) | 42.4 (0.4) |
| 19.7 | 18.3 (1.4) | 16.5 (3.2) | 16.4 (3.3) | 18 (1.7) | 16.8 (2.9) | 17.3 (2.4) | 17.3 (2.4) | 16.1 (3.6) |
| 17.8 | 14.2 (3.6) | 15.6 (2.3) | 14.7 (3.1) | 14.4 (3.5) | 16.4 (1.4) | 16.9 (0.9) | 16.7 (1.1) | 16.4 (1.4) |
| 47.2 | 43.5 (3.7) | 47.2 (0) | 47.8 (0.6) | 44.3 (2.9) | 48.7 (1.5) | 49.7 (2.5) | 50.1 (2.9) | 48.5 (1.3) |
| 49.9 | 46.9 (3) | 48.3 (1.6) | 49.4 (0.5) | 48.7 (1.2) | 51.4 (1.5) | 47 (2.9) | 50.7 (0.8) | 41.8 (8.1) |
| MAE | 2.23 | 1.68 | 1.82 | 1.89 | 2.01 | 1.95 | 1.74 | 2.7 |

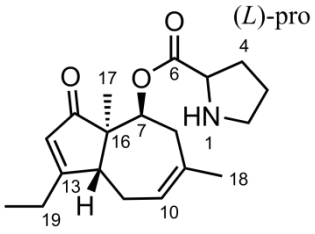
## A4.1.5. Aculenes

**Table A4-13. NMR assignment $^{1}$H of the azumamides**

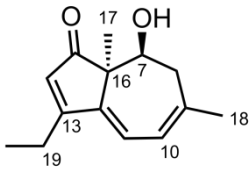
| | **Aculene A** | | | **Aculene B** | | | **Aculene C** | |
|---|---|---|---|---|---|---|---|---|
| # | $\delta_H$ [ppm] (int,mult,*J* [Hz]) | $\delta_C$ [ppm] | # | $\delta_H$ [ppm] (int,mult,*J* [Hz]) | $\delta_C$ [ppm] | # | $\delta_H$ [ppm] (int,mult,*J* [Hz]) | $\delta_C$ [ppm] |
| 1 | - | - | 1 | - | - | 1 | - | - |
| 2 | 3.15 (m) | 45.1 | 2 | 3.21 (m) | 44.7 | 2 | - | - |
| 2’ | 3.09 (m) | 45.1 | 2’ | 3.14 (m) | 44.7 | 2’ | - | - |
| 3 | 1.83 (m) | 22.3 | 3 | 1.84 (m) | 22.3 | 3 | - | - |
| 3’ | 1.67 (m) | 22.3 | 3’ | 1.76 (m) | 22.3 | 3’ | - | - |
| 4 | 1.94 (m) | 27.6 | 4 | 2.05 (m) | 27.6 | 4 | - | - |
| 4’ | 1.46 (m) | 27.6 | 4’ | 1.55 (m) | 27.6 | 4’ | - | - |
| 5 | 4.23 (t 7.7) | 58.4 | 5 | 4.31 (dd 2.4, 5.5) | 58.3 | 5 | - | - |
| 6 | - | 168.0 | 6 | - | 167.8 | 6 | - | - |
| 7 | 5.33 (dd 4.4, 2.4) | 72.9 | 7 | 5.28 (dd 5.3, 2.3) | 73.5 | 7 | 3.95 (t 3.5) | 67.6 |
| 8 | 2.82 (d 20.8) | 36.5 | 8 | 2.67 (d 19.1) | 35.5 | 8 | 2.52 (br. d 20.1) | 40.2 |
| 8’ | 2.55 (m) | 36.5 | 8’ | 2.37 (m) | 35.5 | 8’ | 2.31 (m) | 40.2 |
| 9 | - | 141.7 | 9 | - | 130.7 | 9 | - | 142.5 |
| 10 | 5.98 (d 7.4) | 120.0 | 10 | 5.56 (d 5.5) | 122.7 | 10 | 5.87 (d 7.3) | 119.7 |
| 11 | 6.24 (d 7.4) | 118.5 | 11 | 2.12 (t 15.2) | 24.9 | 11 | 6.08 (d 7.4) | 117.8 |
| 11’ | - | 118.5 | 11’ | 2.52 (m) | 24.9 | 11’ | - | 117.8 |
| 12 | - | 142.5 | 12 | 3.31 (m) | 44.6 | 12 | - | 144.3 |
| 13 | - | 175.3 | 13 | - | 183.8 | 13 | - | 174.3 |
| 14 | 6.03 (s) | 125.3 | 14 | 5.81 (q 1.8) | 122.8 | 14 | 5.93 (s) | 125.6 |
| 15 | - | 205.8 | 15 | - | 208.2 | 15 | - | 208.2 |
| 16 | - | 51.7 | 16 | - | 54.4 | 16 | - | 53.6 |
| 17 | 0.96 (s) | 17.9 | 17 | 0.96 (s) | 17.1 | 17 | 0.79 (s) | 18.3 |
| 18 | 1.85 (s) | 26.9 | 18 | 1.68 (s) | 28.3 | 18 | 1.82 (s) | 27.1 |
| 19 | 2.551 (m) | 20.1 | 19 | 2.39 (m) | 23.1 | 19 | 2.48 (m) | 19.8 |
| 20 | 1.16 (t 7.4) | 12.0 | 20 | 1.11 (t 7.3) | 10.8 | 20 | 1.14 (t 7.4) | 11.8 |
| - | - | - | - | - | - | - | 4.43 (s) | - |
| OH | | | OH | | | OH | | |




Aculene A




Aculene B




Aculene C

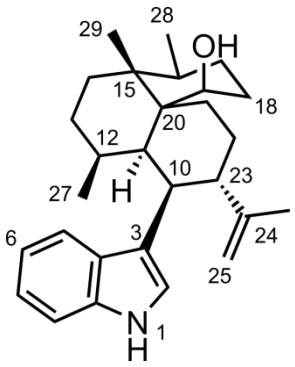
## A4.1.6. Epi-10,23-dihydro-24,25-dehydroaflavine

**Table A4-14. NMR assignment $^{1}$H of the azumamides**

**Table A4-15. Experimental NOE intensities and *J*-coupling constants of the azumamides ($\tau_m$ = 150 ms)**

| # | $\delta_H$ [ppm] (int,mult,*J* [Hz]) | $\delta_C$ [ppm] |
|---|---|---|
| 1 | 10.66 | - |
| 2 | 7.07 (d 2.0) | 122.7 |
| 3 | - | 114.4 |
| 4 | - | 126.7 |
| 5 | 7.38 (d 7.8) | 116.8 |
| 6 | 6.96 (t 7.2) | 117.7 |
| 7 | 7.02 (t 7.3) | 120.0 |
| 8 | 7.29 (d 8.0) | 110.9 |
| 9 | - | 135.5 |
| 10 | 3.59 (dd 13.3, 5.0) | 33.4 |
| 11 | 2.47 (m) | 37.6 |
| 12 | 1.27 (m) | 28.9 |
| 13 | 1.53 (m) | 28.1 |
| 13’ | 0.81 (d 12.9) | 28.1 |
| 14 | 1.46 (m) | 27.4 |
| 14’ | 1.08 (d 13.7) | 27.4 |
| 15 | - | 38.2 |
| 16 | 2.04 (m) | 30.4 |
| 17 | 1.71 (m) | 24.8 |
| 17’ | 1.22 (m) | 24.8 |
| 18 | 1.99 (d 11.3) | 29.6 |
| 18’ | 1.74 (m) | 29.6 |
| 19 | 4.64 (s) | 65.6 |
| 20 | - | 42.9 |
| 21 | 2.07 (m) | 23.5 |
| 21’ | 1.65 (m) | 23.5 |
| 22 | 1.85 (m) | 26.6 |
| 22’ | 1.54 (m) | 26.6 |
| 23 | 3.13 (m) | 42.3 |
| 24 | - | 149.8 |
| 25 | 4.81 (d 1.8) | 110.4 |
| 25’ | 4.58 (d 1.8) | 110.4 |
| 26 | 1.45 (s) | 17.7 |
| 27 | 1.21 (s) | 21.1 |
| 28 | 0.71 (d 6.7) | 15.4 |
| 29 | 0.92 (s) | 17.7 |
| 30 | 4.28 (d 4.3) | - |

| H1 | H2 | $\eta^{norm}$ | Rel |
|---|---|---|---|
| 1 | 2 | 0.0047 | 0.08 |
| 1 | 8 | 0.0021 | 0.04 |
| 2 | 23 | 0.0054 | 0.10 |
| 2 | 25 | 0.0021 | 0.04 |
| 2 | 27 | 0.0080 | 0.14 |
| 5 | 10 | 0.0059 | 0.11 |
| 5 | 11 | 0.0081 | 0.14 |
| 5 | 18 | 0.0039 | 0.07 |
| 10 | 11 | 0.0126 | 0.22 |
| 10 | 19 | 0.0294 | 0.52 |
| 10 | 25 | 0.0034 | 0.06 |
| 10 | 26 | 0.0115 | 0.20 |
| 11 | 12 | 0.0143 | 0.25 |
| 11 | 13 | 0.0042 | 0.07 |
| 11 | 16/18 | 0.0380 | 0.67 |
| 11 | 16/18 | 0.0380 | 0.67 |
| 11 | 19 | 0.0011 | 0.02 |
| 13 | 13’ | 0.0563 | 1.00 |
| 14’ | 28 | 0.0178 | 0.32 |
| 16 | 13 | 0.0313 | 0.56 |
| 16 | 28 | 0.0128 | 0.23 |
| 16/18 | 17’ | 0.0184 | 0.33 |
| 16/18 | 17’ | 0.0184 | 0.33 |
| 17 | 17’ | 0.0608 | 1.08 |
| 17 | 29 | 0.0105 | 0.19 |
| 17’ | 28 | 0.0094 | 0.17 |
| 18 | 19 | 0.0101 | 0.18 |
| 18’ | 19 | 0.0096 | 0.17 |
| 18’ | 30 | 0.0097 | 0.17 |
| 19 | 22 | 0.0233 | 0.41 |
| 19 | 26 | 0.0023 | 0.04 |
| 21 | 29 | 0.0212 | 0.38 |
| 21 | 30 | 0.0046 | 0.08 |
| 21’ | 27 | 0.0160 | 0.28 |
| 22 | 22’ | 0.0551 | 0.98 |
| 22 | 30 | 0.0043 | 0.08 |
| 23 | 25 | 0.0086 | 0.15 |
| 23 | 27 | 0.0240 | 0.43 |
| 25 | 25’ | 0.0400 | 0.71 |
| 25’ | 26 | 0.0088 | 0.16 |
| 29 | 30 | 0.0028 | 0.05 |

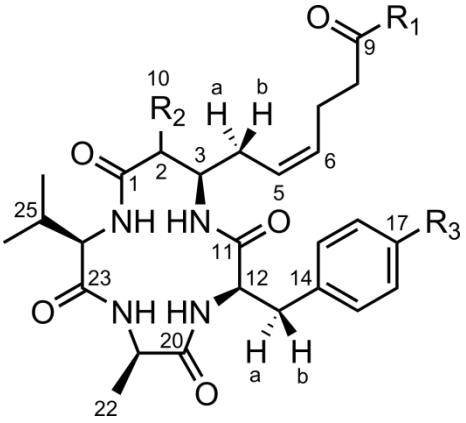


## A4.2. Synthetic peptides

### A4.2.1. Azumamides

**Table A4-16. NMR assignment $^1$H of the azumamides**

| Azu A | $\delta_H$ (int,mult,*J*) | Des C | $\delta_H$ (int,mult,*J*) | Epi E | $\delta_H$ (int,mult,*J*) |
|---|---|---|---|---|---|
| N1 | 7.10 (1H,d,8.0) | N1 | 7.24 (1H,d,7.1) | N1 | 7.35 (1H,d,9.5) |
| N2 | 7.71 (1H,d,9.1) | N2 | 7.54 (1H,d,9.6) | N2 | 7.69 (1H,d,10.9) |
| N3 | 7.68 (1H,d,8.7) | N3 | 7.41 (1H,d,9.7) | N3 | 7.42 (1H,d,11.6) |
| N4 | 7.41 (1H,d,8.2) | N4 | 7.85 (1H,d,6.8) | N4 | 7.87 (1H,d,9.4) |
| 1 | - | 1 | - | 1 | - |
| 2 | 2.53 (1H,qd,7.4,3.8) | 2a | 2.28 (11.3,3.8) | 2 | 2.63 (1H, dq,13.0,6.6) |
| | | 2b | 2.46 (14.0,12.4) | | |
| 3 | 4.00 (1H,qd,7.4,3.7) | 3 | 3.92 (1H,m) | 3 | 3.73 (1H, m) |
| 4a | 2.21 (1H,14.5,8.1) | 4a | 2.25 (2H,m) | 4a | 2.24 (1H,m) |
| 4b | 2.40 (1H,dt,14.2,6.4) | 4b | 2.25 (2H,m) | 4b | 2.40 (1H,dt,14.5,4.6) |
| 5 | 5.27 (1H,m) | 5 | 5.36 (1H,m) | 5 | 5.48 (1H,m) |
| 6 | 5.40 (1H,m) | 6 | 5.47 (1H,m) | 6 | 5.36 (1H,m) |
| 7 | 2.26 (2H,m) | 7 | 2.25 (2H,m) | 7 | 2.26 (2H,m) |
| 8 | 2.09 (2H,m) | 8 | 2.25 (2H,m) | 8 | 2.09 (2H,m) |
| 9 | - | 9 | - | 9 | - |
| 10 | 1.23 (3H,d,7.5) | | | 10 | 1.01 (3H,d,6.8) |
| 11 | - | 11 | - | 11 | - |
| 12 | 4.14 (1H, m) | 12 | 4.15 (1H,td,9.4,5.9) | 12 | 4.26 (1H,m) |
| 13a | 2.93 (1H,13.6,9.4) | 13a | 2.68 (1H,dd,13.9,9.3) | 13a | 2.76 (1H,dd,13.9,10.5) |
| 13b | 3.01 (1H,13.7,6.7) | 13b | 2.81 (1H,dd,13.9,5.7) | 13b | 2.93 (1H,dd,13.9,5.4) |
| 14 | - | 14 | - | 14 | - |
| 15 | 7.25 (2H,d,7.2) | 15 | 6.93 (2H,d,8.4) | 15 | 7.14 (1H,d,7.3) |
| 16 | 7.18 (2H,m) | 16 | 6.63 (2H,d,8.4) | 16 | 7.18 (1H,dd,7.7,7.0) |
| 17 | 7.19 (1H,m) | 17 | - | 17 | 7.25 (1H,t,7.1) |
| 20 | - | 20 | - | 21 | - |
| 21 | 4.13 (1H, m) | 21 | 4.26 (1H,dq,9.7,7.4) | 21 | 4.26 (1H,m) |
| 22 | 1.16 (3H,d,7.3) | 22 | 1.15 (3H,d,7.3) | 22 | 1.13 (3H,d,7.4) |
| 23 | - | 23 | - | 23 | - |
| 24 | 3.73 (1H, dd,10.1,9.2) | 24 | 3.79 (1H,t, 6.6) | 24 | 3.76 (1H,dd,9.4,6.5) |
| 25 | 2.11 (1H,m) | 25 | 1.93 (1H,m) | 25 | 1.96 (1H,m) |
| 26 | 0.86 (3H,d,6.3) | 26 | 0.90 (3H,d,6.8) | 26 | 0.90 (3H,d,6.8) |
| 27 | 0.88 (3H,d,6.3) | 27 | 0.92 (3H,d,6.8) | 27 | 0.92 (3H,d,6.8) |
| CONH1 | 6.73 (1H,br. s) | OH | 9.22 (1H,s) | | |
| CONH2 | 7.26 (1H,br. s) | COOH | 12.07 (1H,br. s) | COOH | 12.07 (1H,br. s) |




| | $R_1$ | $R_2$ | $R_3$ |
|---|---|---|---|
| **Azu A:** | $NH_2$ | (*S*)-Me | H |
| **Des C:** | OH | H | OH |
| **Epi E:** | OH | (*R*)-Me | H |

**Table A4-17. Experimental ROE intensities and *J*-coupling constants of the azumamides ($\tau_m$ = 200 ms)**

**Azumamide A**

| H1 | H2 | $\eta^{norm}$ | Rel |
|---|---|---|---|
| N1 | N4 | 0.0044 | 0.04 |
| N1 | 3 | 0.0124 | 0.11 |
| N1 | 4a | 0.0043 | 0.04 |
| N1 | 4b | 0.0049 | 0.04 |
| N1 | 10 | 0.0277 | 0.25 |
| N1 | 13a | 0.0169 | 0.15 |
| N1 | 13b | 0.0198 | 0.18 |
| N2 | N1 | 0.0431 | 0.39 |
| N2 | 13a | 0.0117 | 0.11 |
| N2 | 13b | 0.0068 | 0.06 |
| N2 | 21 | 0.0127 | 0.12 |
| N2 | 22 | 0.0230 | 0.21 |
| N3 | N4 | 0.0221 | 0.20 |
| N3 | 21 | 0.0101 | 0.09 |
| N3 | 22 | 0.0248 | 0.23 |
| N3 | 25 | 0.0185 | 0.17 |
| N3 | 26/27 | 0.0125 | 0.11 |
| N4 | 2 | 0.0034 | 0.03 |
| N4 | 10 | 0.0331 | 0.30 |
| N4 | 24 | 0.0060 | 0.05 |
| N4 | 25 | 0.0047 | 0.04 |
| N4 | 26/27 | 0.0117 | 0.11 |
| 2 | 3 | 0.0423 | 0.39 |
| 2 | 4a | 0.0091 | 0.08 |
| 2 | 10 | 0.0395 | 0.36 |
| 4a | 4b | 0.1092 | 1.00 |
| 4a | 10 | 0.0299 | 0.27 |
| 4b | 10 | 0.0064 | 0.06 |
| 12 | 15 | 0.0115 | 0.11 |
| 21 | 22 | 0.0260 | 0.24 |
| 24 | 25 | 0.0057 | 0.05 |
| 24 | 26/27 | 0.0373 | 0.34 |
| 25 | 26/27 | 0.0499 | 0.46 |

| H1 | H2 | *J* |
|---|---|---|
| N1 | 3 | 8.0 |
| N2 | 12 | 9.1 |
| N3 | 21 | 8.7 |
| N4 | 24 | 8.3 |
| 3 | 2 | 3.8 |
| 3 | 4a | 8.1 |
| 3 | 4b | 6.4 |
| 12 | 13a | 9.4 |
| 12 | 13b | 6.7 |
| 24 | 25 | 10.1 |

**Des-azumamide C**

| H1 | H2 | $\eta^{norm}$ | Rel |
|---|---|---|---|
| N1 | N2 | 0.0844 | 0.52 |
| N1 | N3 | 0.0113 | 0.07 |
| N1 | 2b | 0.0366 | 0.23 |
| N1 | 3 | 0.0130 | 0.08 |
| N1 | 4a/4b | 0.0118 | 0.07 |
| N1 | 12 | 0.0077 | 0.05 |
| N1 | 13a | 0.0114 | 0.07 |
| N1 | 13b | 0.0075 | 0.05 |
| N2 | N3 | 0.0462 | 0.29 |
| N2 | 13a | 0.0191 | 0.12 |
| N2 | 13b | 0.0091 | 0.06 |
| N2 | 21 | 0.0036 | 0.02 |
| N2 | 22 | 0.0311 | 0.19 |
| N3 | N4 | 0.0356 | 0.22 |
| N3 | 22 | 0.0245 | 0.15 |
| N3 | 24 | 0.0037 | 0.02 |
| N3 | 25 | 0.0092 | 0.06 |
| N3 | 26/27 | 0.0109 | 0.07 |
| N4 | 2b | 0.0776 | 0.48 |
| N4 | 25 | 0.0177 | 0.11 |
| N4 | 26/27 | 0.0379 | 0.23 |
| 2a | 2b | 0.1616 | 1.00 |
| 3 | 2a/4a/4b | 0.0848 | 0.52 |
| 3 | 5 | 0.0079 | 0.05 |
| 6 | 7 | 0.0489 | 0.30 |
| 15 | 12 | 0.0139 | 0.09 |
| 15 | 13a | 0.0263 | 0.16 |
| 15 | 13b | 0.0297 | 0.18 |
| 15 | 22 | 0.0050 | 0.03 |
| 21 | 22 | 0.0511 | 0.32 |
| 24 | 25 | 0.0254 | 0.16 |
| 24 | 26/27 | 0.0420 | 0.26 |
| 25 | 26/27 | 0.0750 | 0.46 |

| H1 | H2 | *J* |
|---|---|---|
| N1 | 3 | 7.3 |
| N2 | 12 | 9.6 |
| N3 | 21 | 9.7 |
| N4 | 24 | 6.8 |
| 2a | 3 | 3.6 |
| 2b | 3 | 12.4 |
| 3 | 4a | 6.5 |
| 3 | 4b | 6.5 |
| 12 | 13a | 9.3 |
| 12 | 13b | 5.8 |
| 24 | 25 | 6.7 |

**Epi-azumamide E**

| H1 | H2 | $\eta^{norm}$ | Rel |
|---|---|---|---|
| N1 | N2 | 0.0945 | 16.28 |
| N1 | 2 | 0.0173 | 2.98 |
| N2 | 22 | 0.0137 | 2.36 |
| N2 | 13a | 0.0068 | 1.17 |
| N3 | N4 | 0.0227 | 3.91 |
| N4 | 2 | 0.0429 | 7.39 |
| N4 | 25 | 0.0060 | 1.03 |
| N4 | 26/27 | 0.0198 | 3.41 |
| 2 | 3 | 0.0058 | 1.00 |
| 2 | 10 | 0.0250 | 4.30 |
| 3 | 10 | 0.0156 | 2.69 |
| 3 | 4a | 0.0259 | 4.47 |
| 3 | 4b | 0.0159 | 2.74 |
| 4a | 4b | 0.0550 | 9.47 |
| 4b | 10 | 0.0231 | 3.98 |
| 6 | 7 | 0.0414 | 7.13 |
| 13b | 13a | 0.0538 | 9.27 |
| 15 | 12 | 0.0068 | 1.17 |
| 15 | 13a | 0.0052 | 0.89 |
| 15 | 13b | 0.0069 | 1.19 |
| 21 | 22 | 0.0364 | 6.27 |
| 24 | 25 | 0.0146 | 2.51 |
| 24 | 26/27 | 0.0317 | 5.47 |
| 25 | 26/27 | 0.0422 | 7.28 |

| H1 | H2 | *J* |
|---|---|---|
| N1 | 3 | 9.5 |
| N2 | 12 | 10.9 |
| N3 | 21 | 11.6 |
| N4 | 24 | 9.4 |
| 2 | 3 | 13.0 |
| 3 | 4b | 5.6 |
| 3 | 4a | 4.6 |
| 12 | 13a | 10.5 |
| 12 | 13b | 5.4 |

**Table A4-18. Experimental distances from relative ROE intensities and *J*-coupling constants compared to best fit of multiple structures for the azumamides ($\tau_m$ = 200 ms)**

**Azumamide A**

| H1 | H2 | NOE | Calc | Diff | % |
|---|---|---|---|---|---|
| N1 | N2 | 2.13 | 2.15 | 0.03 | 1.3 |
| N1 | N4 | 3.11 | 3.11 | 0.00 | 0.0 |
| N1 | 3 | 2.62 | 2.66 | 0.04 | 1.4 |
| N1 | 4a | 3.12 | 3.12 | 0.00 | 0.0 |
| N1 | 4b | 3.06 | 3.06 | 0.00 | 0.0 |
| N1 | 10 | 2.75 | 2.80 | 0.04 | 1.8 |
| N1 | 13a | 2.49 | 2.49 | 0.00 | 0.0 |
| N1 | 13b | 2.42 | 2.44 | 0.02 | 0.7 |
| N2 | 13a | 2.64 | 2.64 | 0.00 | 0.0 |
| N2 | 13b | 2.89 | 2.89 | 0.00 | 0.0 |
| N2 | 21 | 2.61 | 2.86 | 0.25 | 8.8 |
| N2 | 22 | 2.83 | 3.15 | 0.26 | 9.9 |
| N3 | N4 | 2.38 | 2.35 | 0.02 | 1.0 |
| N3 | 22 | 2.80 | 2.97 | 0.14 | 5.6 |
| N3 | 21 | 2.71 | 2.87 | 0.16 | 5.5 |
| N3 | 25 | 2.45 | 2.45 | 0.00 | 0.0 |
| N3 | 26/27 | 3.13 | 3.13 | 0.00 | 0.0 |
| N4 | 2 | 3.24 | 3.33 | 0.08 | 2.5 |
| N4 | 10 | 2.67 | 2.49 | 0.15 | 7.3 |
| N4 | 24 | 2.96 | 2.97 | 0.01 | 0.4 |
| N4 | 25 | 3.07 | 2.68 | 0.39 | 14.7 |
| N4 | 26/27 | 3.17 | 2.97 | 0.17 | 7.0 |
| 2 | 10 | 2.59 | 2.68 | 0.07 | 3.2 |
| 2 | 3 | 2.13 | 2.39 | 0.26 | 10.7 |
| 2 | 4a | 2.76 | 2.76 | 0.00 | 0.0 |
| **4b** | **4a** | **1.82** | **1.82** | **0.00** | **0.0** |
| 4a | 10 | 2.71 | 2.97 | 0.22 | 8.7 |
| 4b | 10 | 3.52 | 3.51 | 0.01 | 0.3 |
| 12 | 15 | 2.65 | 2.65 | 0.00 | 0.0 |
| 13a | 15 | 2.50 | 2.44 | 0.06 | 2.5 |
| 13b | 15 | 2.47 | 2.47 | 0.00 | 0.0 |
| 21 | 22 | 2.77 | 2.77 | 0.00 | 0.1 |
| 24 | 25 | 2.98 | 2.85 | 0.13 | 4.7 |
| 24 | 26/27 | 2.62 | 2.81 | 0.16 | 6.9 |
| 25 | 26/27 | 2.50 | 2.43 | 0.06 | 2.7 |
| | | | MAE | 0.08 | 3.1 |

| H1 | H2 | Exp | Calc | Diff |
|---|---|---|---|---|
| N1 | 3 | 8.0 | 8.0 | 0.0 |
| N2 | 12 | 9.1 | 8.9 | 0.2 |
| N3 | 21 | 8.7 | 8.2 | 0.5 |
| N4 | 24 | 8.3 | 8.3 | 0.0 |
| 2 | 3 | 3.8 | 3.8 | 0.0 |
| 3 | 4a | 8.1 | 8.1 | 0.0 |
| 3 | 4b | 6.4 | 6.4 | 0.0 |
| 12 | 13a | 9.4 | 8.8 | 0.6 |
| 12 | 13b | 6.7 | 6.7 | 0.0 |
| 24 | 25 | 10.1 | 10.1 | 0.0 |

**Des-azumamide C**

| H1 | H2 | NOE | Calc | Diff | % |
|---|---|---|---|---|---|
| N1 | N3 | 2.79 | 2.80 | 0.01 | 0.4 |
| N1 | 2b | 2.29 | 2.49 | 0.19 | 7.8 |
| N1 | 3 | 2.73 | 2.70 | 0.03 | 1.0 |
| N1 | 4a/4b | 3.11 | 3.00 | 0.11 | 3.6 |
| N1 | 12 | 2.98 | 3.11 | 0.14 | 4.5 |
| N1 | 13a | 2.79 | 2.72 | 0.06 | 2.3 |
| N1 | 13b | 2.99 | 3.00 | 0.02 | 0.6 |
| N1 | N2 | 1.99 | 1.91 | 0.08 | 4.4 |
| N2 | N3 | 2.21 | 2.20 | 0.00 | 0.0 |
| N2 | 13a | 2.56 | 2.59 | 0.03 | 1.2 |
| N2 | 13b | 2.89 | 2.90 | 0.01 | 0.3 |
| N2 | 21 | 3.38 | 3.55 | 0.17 | 4.8 |
| N2 | 22 | 2.83 | 2.87 | 0.04 | 1.3 |
| N3 | N4 | 2.30 | 2.37 | 0.06 | 2.7 |
| N3 | 22 | 2.94 | 2.96 | 0.02 | 0.6 |
| N3 | 24 | 3.36 | 3.45 | 0.09 | 2.6 |
| N3 | 25 | 2.88 | 2.99 | 0.11 | 3.6 |
| N3 | 26/27 | 3.78 | 3.79 | 0.01 | 0.2 |
| N4 | 2b | 2.02 | 2.18 | 0.16 | 7.1 |
| N4 | 25 | 2.59 | 2.56 | 0.03 | 1.1 |
| N4 | 26/27 | 3.07 | 3.07 | 0.00 | 0.0 |
| **2b** | **2a** | **1.79** | **1.79** | **0.00** | **0.0** |
| 3 | 2a/4a/4b | 1.99 | 2.07 | 0.08 | 3.8 |
| 5 | 3 | 2.96 | 2.94 | 0.01 | 0.5 |
| 6 | 7 | 2.45 | 2.45 | 0.00 | 0.2 |
| 12 | 15 | 3.02 | 2.86 | 0.16 | 5.7 |
| 13b | 13a | 2.08 | 1.70 | 0.38 | 22.1 |
| 13a | 15 | 2.72 | 2.44 | 0.28 | 11.3 |
| 13b | 15 | 2.66 | 2.47 | 0.20 | 8.0 |
| 15 | 22 | 4.31 | 4.25 | 0.06 | 1.4 |
| 21 | 22 | 2.60 | 2.62 | 0.01 | 0.5 |
| 24 | 25 | 2.44 | 2.56 | 0.12 | 4.8 |
| 24 | 26/27 | 3.02 | 3.19 | 0.17 | 5.4 |
| 25 | 26/27 | 2.74 | 2.59 | 0.15 | 5.7 |
| | | | MAE | 0.09 | 3.5 |

| H1 | H2 | Exp | Calc | Diff |
|---|---|---|---|---|
| N1 | 3 | 7.3 | 7.4 | 0.1 |
| N2 | 12 | 9.6 | 9.1 | 0.4 |
| N3 | 21 | 9.7 | 9.0 | 0.7 |
| N4 | 24 | 6.8 | 6.8 | 0.0 |
| 2b | 3 | 12.4 | 12.4 | 0.0 |
| 2a | 3 | 3.6 | 3.6 | 0.0 |
| 3 | 4a | 6.5 | 6.5 | 0.0 |
| 3 | 4b | 6.5 | 6.6 | 0.1 |
| 12 | 13a | 9.3 | 9.3 | 0.0 |
| 12 | 13b | 5.8 | 5.8 | 0.0 |
| 24 | 25 | 6.7 | 6.8 | 0.1 |

**Epi-azumamide E**

| H1 | H2 | Calc | NOE | Diff | Diff % |
|---|---|---|---|---|---|
| N1 | N2 | 1.88 | 1.91 | 0.02 | 1.1 |
| N1 | 2 | 2.50 | 2.49 | 0.01 | 0.3 |
| N4 | 2 | 2.15 | 2.12 | 0.02 | 1.1 |
| N2 | 13a | 2.92 | 2.92 | 0.00 | 0.1 |
| N2 | 22 | 3.12 | 3.13 | 0.01 | 0.3 |
| N3 | N4 | 2.39 | 2.29 | 0.10 | 4.2 |
| N4 | 25 | 2.98 | 2.99 | 0.00 | 0.1 |
| N4 | 26/27 | 2.93 | 2.75 | 0.151 | 6.6 |
| **2** | **3** | **3.00** | **3.00** | **0.00** | **0.0** |
| 2 | 10 | 2.82 | 2.74 | 0.07 | 3.1 |
| 3 | 4a | 2.34 | 2.35 | 0.01 | 0.5 |
| 3 | 4b | 2.53 | 2.62 | 0.09 | 3.3 |
| 3 | 10 | 3.05 | 3.07 | 0.02 | 0.6 |
| 4a | 4b | 2.06[b] | 1.71 | 0.35 | 20.4 |
| 4b | 10 | 2.86 | 2.90 | 0.03 | 1.4 |
| 6 | 7 | 2.16 | 2.32 | 0.16 | 6.9 |
| 12 | 15 | 2.92 | 2.80 | 0.12 | 4.2 |
| 13b | 13a | 2.07[b] | 1.70 | 0.37 | 21.5 |
| 13a | 15 | 3.05[b] | 2.46 | 0.59 | 23.9 |
| 13b | 15 | 2.91[b] | 2.50 | 0.41 | 16.4 |
| 21 | 22 | 2.65 | 2.78 | 0.11 | 4.6 |
| 24 | 25 | 2.57 | 2.66 | 0.09 | 3.4 |
| 24 | 26/27 | 2.71 | 2.87 | 0.13 | 5.4 |
| 25 | 26/27 | 2.59 | 2.43 | 0.13 | 6.5 |
| | | MAE | | 0.13 | 6.11 |

| H1 | H2 | Calc | Exp | Diff |
|---|---|---|---|---|
| N1 | 3 | 9.5 | 8.7 | 0.8 |
| N2 | 12 | 10.9 | 8.9 | 2 |
| N3 | 21 | 11.6 | 9.1 | 2.5 |
| N4 | 24 | 9.4 | 8.5 | 0.9 |
| 2 | 3 | 13.0 | 12.4 | 0.6 |
| 3 | 4a | 4.6 | 4.6 | 0.0 |
| 3 | 4b | 5.6 | 5.6 | 0.0 |
| 12 | 13a | 10.5 | 10.0 | 0.5 |
| 12 | 13b | 5.4 | 5.2 | 0.2 |

**Azumamide A, constrained**

| H1 | H2 | Exp | Calc | Diff % |
|---|---|---|---|---|
| N1 | 10 | 2.19 | 2.3 | 5.0 |
| N1 | 4a | 2.99 | 3.29 | 10.0 |
| N1 | 4b | 2.92 | 3.34 | 14.4 |
| N1 | 13a | 2.38 | 2.345 | 1.5 |
| N1 | 13b | 2.32 | 2.535 | 9.3 |
| N1 | 3 | 2.51 | 2.89 | 15.1 |
| N2 | 22 | 2.26 | 3.87 | 71.2 |
| N2 | 13a | 2.53 | 2.69 | 6.3 |
| N2 | 13b | 2.76 | 3.535 | 28.1 |
| N2 | 21 | 2.49 | 2.825 | 13.5 |
| N2 | N1 | 2.03 | 2.145 | 5.7 |
| N3 | 26/27 | 2.5 | 2.915 | 16.6 |
| N3 | 22 | 2.23 | 2.7 | 21.1 |
| N3 | 25 | 2.34 | 2.09 | 10.7 |
| N3 | 21 | 2.59 | 2.975 | 14.9 |
| N3 | N4 | 2.27 | 2.675 | 17.8 |
| N4 | 26/27 | 2.53 | 2.755 | 8.9 |
| N4 | 10 | 2.13 | 2.22 | 4.2 |
| N4 | 25 | 2.94 | 2.535 | 13.8 |
| N4 | 2 | 3.1 | 3.555 | 14.7 |
| N4 | 24 | 2.83 | 2.975 | 5.1 |
| N4 | N1 | 2.97 | 2.12 | 28.6 |
| 2 | 10 | 2.06 | 2.28 | 10.7 |
| 2 | 4a | 2.64 | 2.945 | 11.6 |
| 3 | 2 | 2.04 | 2.435 | 19.4 |
| 4a | 10 | 2.16 | 2.325 | 7.6 |
| 4b | 10 | 2.8 | 3.71 | 32.5 |
| **4b** | **4a** | **1.745** | **1.745** | **0.0** |
| 15 | 13a | 2.39 | 3.755 | 57.1 |
| 15 | 13b | 2.37 | 2.615 | 10.3 |
| 15 | 12 | 2.54 | 2.865 | 12.8 |
| 21 | 22 | 2.21 | 2.34 | 5.9 |
| 24 | 25 | 2.85 | 3.075 | 7.9 |
| 24 | 26/27 | 2.08 | 2.31 | 11.1 |
| 25 | 26/27 | 1.98 | 2.065 | 4.3 |

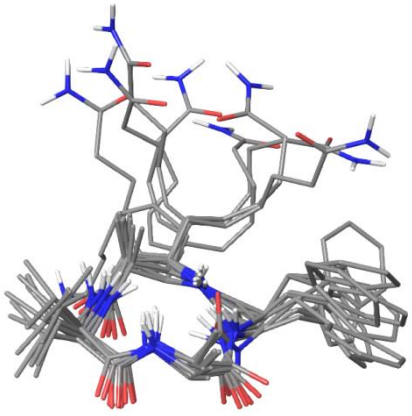


**Figure A4.3. Representative structures of the conformational space covered by azumamide A.**

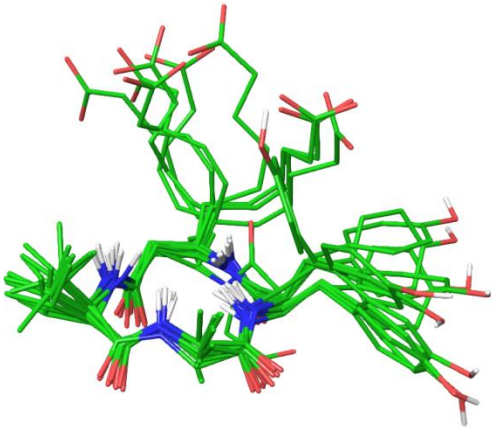


**Figure A4.4. Representative structures of the conformational space covered by des-azumamide C.**

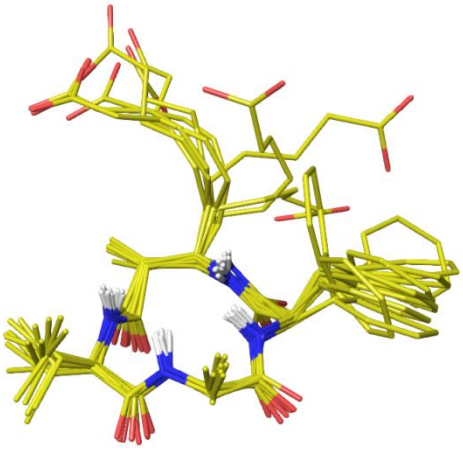


**Figure A4.5. Representative structures of the conformational space covered by epi-azumamide E.**

## A4.2.2. Molecular recognition

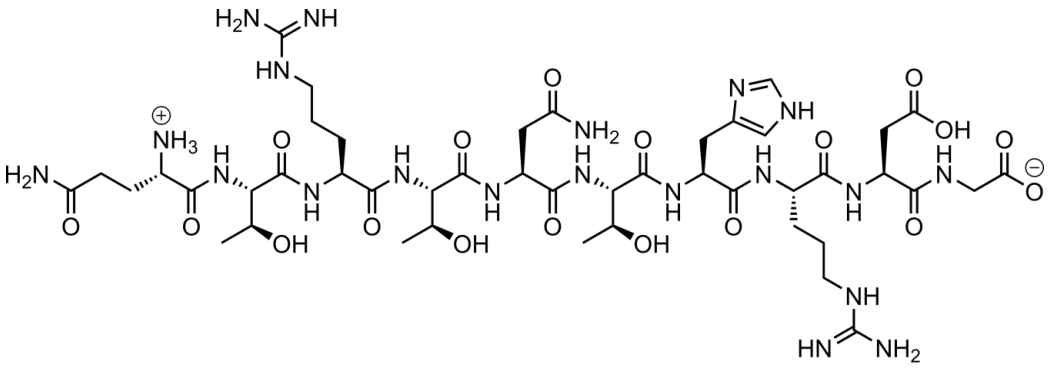
**Table A4-19. NMR assignment of Target 20 and Ligand 7**

| Target 20 | | |
|---|---|---|
| **Name** | **$\delta_H$ [ppm] (int,mult,*J* [Hz])** | **$\delta_C$ [ppm]** |
| Gln(1)-NH | - | - |
| Gln(1)-αH | 4.05 (1H,m) | 52.4 |
| Gln(1)-βH | 2.06 (2H,m) | 26.5 |
| Gln(1)-γH | 2.35 (2H,m) | 29.9 |
| Thr(2)-NH | 8.71 (1H,m) | - |
| Thr(2)-αH | 4.27 (1H,broad s) | 59.4 |
| Thr(2)-βH | 4.06 (1H,m) | 67.0 |
| Thr(2)-γH2 | 1.13 (1H,d,6.3) | 18.7 |
| Arg(3)-NH | 8.55 (1H,d,7.0) | - |
| Arg(3)-αH | 4.35 (1H, q, 7.4) | 53.4 |
| Arg(3)-βH2 | 1.69 (1H,m) | 28.0 |
| Arg(3)-βH3 | 1.78 (1H,m) | 28.0 |
| Arg(3)-γH2 | 1.57 (1H,m) | 24.2 |
| Arg(3)-γH3 | 1.52 (1H,m) | 24.2 |
| Arg(3)- δH | 3.11 (1H,m) | 40.4 |
| Arg(3)-ε/ηH | 7.13 (1H,m) | - |
| Thr(4)-NH | 8.22 (1H,d,7.7) | - |
| Thr(4)-αH | 4.23 (1H,m) | 59.1/53.4 |
| Thr(4)-βH2 | 4.09 (1H,m) | 67.2 |
| Thr(4)-γH | 1.08 (1H,d,6.3) | 18.6 |
| Asn(5)-NH | 8.48 (1H,m) | - |
| Asn(5)-αH | 4.69 (1H,m) | 50.3 |
| Asn(5)-βH2 | 2.77 (1H,dd,16.5,7.0) | 35.8 |
| Asn(5)-βH3 | 2.68 (1H,m) | 35.8 |
| Thr(6)-NH | 8.06 (1H,d,7.6) | - |
| Thr(6)-αH | 4.17 (1H,m) | 59.2 |
| Thr(6)-βH | 4.10 (1H,m) | 66.8 |
| Thr(6)-γH2 | 1.05 (1H,d,6.7) | 18.6 |
| His(7)-NH | 8.43 (1H,d,7.6) | - |
| His(7)-αH | 4.56 (1H,m) | 52.8/51.5 |
| His(7)-βH2 | 3.18 (1H,dd,15.2,6.5) | 25.9 |
| His(7)-βH3 | 3.10 (1H,m) | 25.9 |
| His(7)-δH | 7.17 (1H,s) | 117.5 |
| His(7)-εH | 8.49 (1H,m) | 133.7 |
| Arg(8)-NH | 8.29 (1H,d,8.2) | - |
| Arg(8)-αH | 4.23 (1H,m) | 59.1/53.4 |
| Arg(8)-βH2 | 1.73 (1H,m) | 28.1 |
| Arg(8)-βH3 | 1.63 (1H,m) | 28.1 |
| Arg(8)-γH2 | 1.52 (1H,m) | 24.1 |
| Arg(8)-γH3 | 1.49 (1H,m) | 24.1 |
| Arg(8)- δH | 3.09 (1H,m) | 40.4 |
| Arg(8)-ε/ηH | 7.12 (1H,m) | - |
| Asp(9)-NH | 8.40 (1H,d,7.6) | - |
| Asp(9)-αH | 4.56 (1H,m) | 52.8/51.5 |
| Asp(9)-βH2 | 2.51 (1H,dd,15.9,8.9) | 38.3 |
| Asp(9)-βH3 | 2.66 (1H,m) | 38.3 |
| Gly(10)-NH | 7.82 (1H,t,5.5) | - |
| Gly(10)-αH2 | 3.63 (1H,dd,17.2,5.7) | 43.4 |
| Gly(10)-αH3 | 3.70 (1H,dd,17.0,5.8) | 43.4 |

| Ligand 7 | | |
|---|---|---|
| **Name** | **$\delta_H$ [ppm] (int,mult,*J* [Hz])** | **$\delta_C$ [ppm]** |
| Glu(1)-NH | - | - |
| Glu(1)-αH | 3.85 (1H, broad m) | 52.6 |
| Glu(1)- βH | 1.88 (2H,m) | 26.7 |
| Glu(1)-γH | 2.19 (2H,m) | 31.5 |
| Asp(2)-NH | 8.67 (1H,d,6.7) | - |
| Asp(2)-αH | 4.53 (1H, broad m) | 51.0 |
| Asp(2)-βH2 | 2.59 (1H,dd,17.0,3.5) | 37.7 |
| Asp(2)-βH3 | 2.46 (1H,dd,16.9,9.4) | 37.7 |
| Tyr(3)-NH | 8.17 (1H,m) | - |
| Tyr(3)-αH | 4.43 (1H, broad m) | 55.0 |
| Tyr(3)-βH2 | 2.94 (1H,dd,13.5,5.1) | 35.8 |
| Tyr(3)-βH3 | 2.81 (1H,dd,13.3,8.9) | 35.8 |
| Tyr(3)-γH | 6.99 (2H,d,7.5) | 130.5 |
| Tyr(3)-εH | 6.69 (2H,d,7.5) | 115.5 |
| Glu(4)-NH | 8.14 (1H,m) | - |
| Glu(4)-αH | 4.18 (1H, broad m) | 53.0 |
| Glu(4)-βH2 | 1.81 (1H,m) | 26.6 |
| Glu(4)-βH3 | 1.92 (1H,m) | 26.6 |
| Glu(4)-γH | 2.19 (1H,m) | 31.4 |
| Val(5)-NH | 7.99 (1H,m) | - |
| Val(5)-αH | 3.91 (1H, broad m) | 59.5 |
| Val(5)-βH | 1.93 (1H,m) | 29.9 |
| Val(5)-γH2 | 0.82 (3H,d,6.3) | 18.1 |
| Val(5)-γH3 | 0.80 (3H,d,6.8) | 18.1 |
| Glu(6)-NH | 8.32 (1H,m) | - |
| Glu(6)-αH | 4.24 (1H, broad m) | 53.1 |
| Glu(6)-βH2 | 1.84 (1H,m) | 26.4 |
| Glu(6)-βH3 | 1.97 (1H,m) | 26.4 |
| Glu(6)-γH | 2.27 (1H,m) | 31.2 |
| Glu(7)-NH | 8.32 (1H,m) | - |
| Glu(7)-αH | 4.24 (1H, broad m) | 53.1 |
| Glu(7)-βH2 | 1.84 (1H,m) | 26.4 |
| Glu(7)-βH3 | 1.97 (1H,m) | 26.4 |
| Glu(6)-γH | 2.27 (1H,m) | 31.2 |
| Gly(8)-NH | 8.00 (1H,m) | - |
| Gly(8)-αH | 3.65 (2H, m) | 42.9 |

**Table A4-20. Experimental distances from relative NOE intensities for Target 20 and Ligand 7 as individual peptides ($\tau_m$ = 200 ms).**

| **Target** | | | **Ligand** | | |
|---|---|---|---|---|---|
| **Nuc 1** | **Nuc2** | **Dist (Å)** | **Nuc 1** | **Nuc2** | **Dist (Å)** |
| Thr(2)-NH | Gln(1)-αH | 3.21 | Asp(2)-NH | Glu(1)-αH | 2.44 |
| Arg(3)-NH | Thr(2)-βH | 3.13 | Asp(2)-NH | Asp(2)-βH3 | 2.86 |
| Arg(3)-NH | Thr(2)-αH | 2.42 | Asp(2)-NH | Asp(2)-βH2 | 3.30 |
| Asn(5)-NH | Thr(4)-αH | 2.43 | Tyr(3)-NH | Asp(2)-αH | 2.35 |
| Asn(5)-NH | Thr(4)-βH2 | 2.83 | Tyr(3)-NH | Tyr(3)-βH3 | 2.88 |
| His(7)-NH | Thr(6)-αH | 2.30 | Tyr(3)-NH | Tyr(3)-βH2 | 2.94 |
| His(7)-NH | Thr(6)-βH | 2.98 | Glu(4)-NH | Tyr(3)-αH | 2.32 |
| Asp(9)-NH | Arg(8)-αH | 2.46 | Glu(4)-NH | Glu(4)-βH2 | 2.99 |
| Arg(8)-NH | His(7)-αH | 2.16 | Glu(4)-NH | Glu(4)-βH3 | 3.53 |
| Thr(4)-NH | Thr(4)-βH2 | 2.85 | Glu(4)-NH | Val(5)-NH | 2.96 |
| Thr(4)-NH | Arg(3)-αH | 2.38 | Val(5)-NH | Glu(4)-αH | 2.51 |
| Thr(6)-NH | Thr(6)-βH | 2.81 | Val(5)-NH | Val(5)-βH | 2.65 |
| Gly(10)-NH | Asp(9)-αH | 3.19 | Val(5)-NH | Val(5)-γH2 | 3.05 |
| His(7)-NH | His(7)-βH3 | 2.82 | Glu(6)-NH | Val(5)-αH | 2.33 |
| His(7)-NH | His(7)-βH2 | 2.77 | Glu(6)-NH | Val(5)-NH | 3.04 |
| His(7)-NH | Asp(9)-βH2 | 3.42 | Glu(7)-NH | Glu(7)-βH2 | 2.67 |
| Asn(5)-NH | Asn(5)-βH3 | 2.94 | Gly(8)-NH | Glu(7)-αH | 2.88 |
| Asn(5)-NH | Asn(5)-βH2 | 2.98 | Gly(8)-NH | Glu(7)-NH | 2.99 |
| Arg(3)-NH | Arg(3)-βH2 | 2.79 | Tyr(3)-αH | Tyr(3)-βH2 | 2.54 |
| Arg(3)-NH | Arg(3)-βH3 | 3.09 | Tyr(3)-γH | Tyr(3)-βH3 | 3.14 |
| Asp(9)-NH | Asp(9)-βH2 | 3.03 | Tyr(3)-γH | Tyr(3)-βH2 | 3.22 |
| Asp(9)-NH | Asp(9)-βH3 | 3.75 | **Tyr(3)-βH2** | **Tyr(3)-βH3** | **1.78** |
| Arg(8)-NH | Arg(8)-βH3 | 2.81 | | | |
| Arg(8)-NH | Arg(8)-βH2 | 3.05 | | | |
| His(7)-NH | Thr(6)-NH | 2.55 | | | |
| Arg(3)-NH | Thr(4)-NH | 3.29 | | | |
| Thr(2)-αH | Thr(2)-γH2 | 4.04 | | | |
| Thr(4)-αH | Thr(4)-γH | 3.36 | | | |
| Thr(6)-αH | Thr(6)-γH2 | 3.20 | | | |
| **Arg(8)-βH2** | **Arg(8)-βH3** | **1.78** | | | |



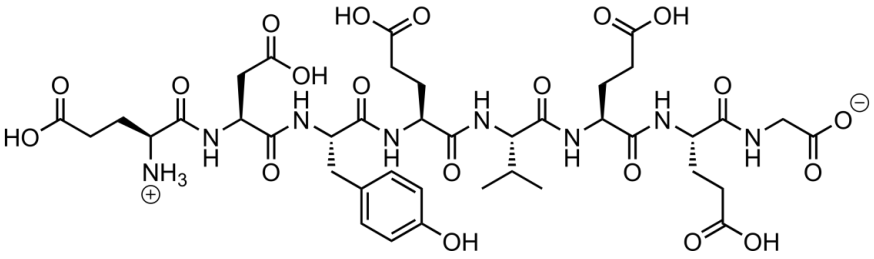


**Table A4-21. Experimental distances from relative NOE intensities for Target 20 and Ligand 7 upon mixing the peptides ($\tau_m$ = 200 ms)**

| **Target 20** | | | **Ligand 7** | | |
|---|---|---|---|---|---|
| **H1** | **H2** | **Dist (Å)** | **H1** | **H2** | **Dist (Å)** |
| Thr(2)-NH | Gln(1)-αH | 2.94 | Asp(2)-NH | Glu(1)-αH | 2.44 |
| Arg(3)-NH | Thr(2)-βH | 2.95 | Asp(2)-NH | Asp(2)-βH3 | 3.16 |
| Arg(3)-NH | Thr(2)-αH | 2.50 | Asp(2)-NH | Asp(2)-βH2 | 3.02 |
| Arg(3)-NH | Arg(3)-δH | 3.60 | Tyr(3)-NH | Asp(2)-αH | 2.55 |
| Arg(3)-NH | Thr(2)-γH2 | 3.36 | Tyr(3)-NH | Tyr(3)-βH2 | 2.94 |
| Arg(3)-NH | Arg(3)-γH | 2.64 | Tyr(3)-NH | Tyr(3)-βH3 | 2.77 |
| Arg(3)-NH | Arg(3)-βH2 | 2.72 | Tyr(3)-NH | Glu(1)-βH2/Asp(2)-βH2 | 3.26 |
| Arg(3)-NH | Arg(3)-βH3 | 3.28 | Tyr(3)-NH | Tyr(3)-γH | 3.64 |
| Thr(4)-NH | Thr(4)-βH2 | 2.94 | Glu(4)-NH | Tyr(3)-αH | 2.31 |
| Thr(4)-NH | Arg(3)-αH | 2.54 | Glu(4)-NH | Glu(4)-βH3 | 3.45 |
| Thr(4)-NH | Thr(4)-γH | 3.31 | Glu(4)-NH | Glu(1)-γH/Glu(4)-γH | 3.49 |
| Thr(4)-NH | Arg(3)-γH | 3.08 | Glu(4)-NH | Tyr(3)-βH3 | 3.24 |
| Thr(4)-NH | Arg(3)-βH2 | 3.34 | Glu(4)-NH | Tyr(3)-βH2 | 3.26 |
| Asn(5)-NH | Thr(4)-αH | 2.37 | Glu(4)-NH | Val(5)-NH | 2.98 |
| Asn(5)-NH | Thr(4)-γH | 3.28 | Val(5)-NH | Glu(4)-αH | 2.18 |
| Asn(5)-NH | His(7)-δH | 3.37 | Val(5)-NH | Tyr(3)-αH | 3.84 |
| Thr(6)-NH | Thr(6)-βH | 2.67 | Val(5)-NH | Val(5)-γH2 | 3.47 |
| Thr(6)-NH | Thr(6)-γH2 | 2.99 | Val(5)-NH | Glu(4)-βH2 | 3.52 |
| Thr(6)-NH | Asn(5)-βH2 | 3.21 | Val(5)-NH | Glu(4)-βH3 | 2.97 |
| His(7)-NH | Thr(6)-αH | 2.22 | Glu(6)-NH | Val(5)-αH | 2.25 |
| His(7)-NH | Thr(6)-βH | 2.65 | Glu(6)-NH | Val(5)-γH2 | 3.71 |
| His(7)-NH | His(7)-βH2 | 2.59 | Glu(6)-NH | Val(5)-βH | 3.03 |
| His(7)-NH | Thr(6)-NH | 2.38 | Glu(6)-NH | Glu(6)-βH3 | 2.67 |
| His(7)-NH | Thr(6)-γH2 | 3.24 | Glu(6)-NH | Val(5)-NH | 2.89 |
| Arg(8)-NH | His(7)-αH | 2.31 | Gly(8)-NH | Glu(7)-αH | 2.66 |
| Arg(8)-NH | Arg(8)-βH3 | 2.60 | Tyr(3)-αH | Tyr(3)-βH2 | 2.64 |
| Arg(8)-NH | Arg(8)-βH2 | 2.44 | Val(5)-αH | Val(5)-γH2 | 3.30 |
| Arg(8)-NH | His(7)-βH2 | 2.91 | Val(5)-αH | Glu(4)-βH3 | 3.26 |
| Arg(8)-NH | His(7)-βH3 | 2.81 | Glu(4)-βH3 | Val(5)-γH2 | 3.03 |
| Arg(8)-NH | Arg(8)-γH3 | 2.73 | Val(5)-βH | Glu(6)-βH3 | 2.28 |
| Asp(9)-NH | Arg(8)-αH | 2.21 | Glu(6)-βH3 | Val(5)-βH | 2.64 |
| Asp(9)-NH | Asp(9)-βH2 | 2.53 | Glu(6)-βH3 | Val(5)-βH | 2.39 |
| Asp(9)-NH | Asp(9)-βH3 | 2.70 | Asp(2)-NH | Glu(1)-αH | 2.63 |
| Asp(9)-NH | Arg(8)-γH3 | 2.96 | Asp(2)-NH | Asp(2)-βH3 | 2.66 |
| Asp(9)-NH | Arg(8)-βH3 | 2.85 | **Tyr(3)-βH2** | **Tyr(3)-βH3** | **1.78** |
| Asp(9)-NH | Arg(8)-βH2 | 2.76 | | | |
| Asp(9)-NH | Gly(10)-NH | 2.74 | | | |
| Gly(10)-NH | Asp(9)-αH | 2.60 | | | |
| Gly(10)-NH | Asp(9)-βH2 | 3.22 | | | |
| Gly(10)-NH | Asp(9)-βH3 | 3.10 | | | |
| Arg(3)-αH | Arg(3)-βH2 | 2.98 | | | |
| Thr(4)-αH | Thr(4)-γH | 2.78 | | | |
| Thr(6)-αH | Thr(6)-γH2 | 2.64 | | | |
| Arg(8)-αH | His(7)-βH3 | 2.46 | | | |
| Asp(9)-αH | Asp(9)-βH3 | 2.43 | | | |
| His(7)-δH | Arg(8)-δH | 2.84 | | | |
| His(7)-δH | Arg(8)-αH | 3.86 | | | |
| His(7)-δH | His(7)-αH | 3.36 | | | |
| His(7)-εH | Thr(6)-βH | 2.99 | | | |
| His(7)-εH | Asn(5)-βH3 | 3.15 | | | |
| His(7)-εH | Asn(5)-βH2 | 3.29 | | | |
| His(7)-βH3 | Arg(8)-βH3 | 2.54 | | | |
| Arg(8)-δH | Arg(8)-βH2 | 2.48 | | | |
| **Arg(8)-βH2** | **Arg(8)-βH3** | **1.78** | | | |

**Table A4-22. Differences in chemical shifts of Target 20 observed upon mixing the peptide Target 20 and Ligand 7**

| Target 20 | $\delta_H$ | Δδ [Hz] | Δδ [ppm] | | $\delta_H$ | Δδ [Hz] | Δδ [ppm] |
|---|---|---|---|---|---|---|---|
| Gln(1)-NH | - | | | Thr(6)-NH | 8.06 | 2.7 | 0 |
| Gln(1)-αH | 4.05 | 11.4 | 0 | Thr(6)-αH | 4.17 | 6.3 | 0.01 |
| Gln(1)-βH | 2.06 | 13.1 | 0.02 | Thr(6)-βH | 4.10 | 8.0 | 0 |
| Gln(1)-γH | 2.35 | | | Thr(6)-γH2 | 1.05 | 6.8 | 0.01 |
| Thr(2)-NH | 8.71 | 0 | 0 | His(7)-NH | 8.43 | 4.4 | 0 |
| Thr(2)-αH | 4.27 | | | His(7)-αH | 4.56 | 10.8 | 0.01 |
| Thr(2)-βH | 4.06 | 11.4 | 0.01 | His(7)-βH2 | 3.18 | 9.7 | 0.01 |
| Thr(2)-γH2 | 1.13 | 11.4 | 0.01 | His(7)-βH3 | 3.10 | 9.1 | 0.01 |
| Arg(3)-NH | 8.55 | 0 | 0 | His(7)-δH | 7.17 | 10.5 | 0.01 |
| Arg(3)-αH | 4.35 | 9.1 | 0.01 | His(7)-εH | 8.49 | 11.7 | 0.01 |
| Arg(3)-βH2 | 1.69 | | | Arg(8)-NH | 8.29 | 2.4 | 0 |
| Arg(3)-βH3 | 1.78 | | | Arg(8)-αH | 4.23 | 2.1 | 0 |
| Arg(3)-γH | 1.57 | | | Arg(8)-βH2 | 1.73 | | |
| Arg(3)-γH | 1.52 | | | Arg(8)-βH3 | 1.63 | | |
| Arg(3)-δH | 3.11 | | | Arg(8)-γH2 | 1.52 | | |
| Arg(3)-δH | 7.13 | | | Arg(8)-γH3 | 1.49 | | |
| Thr(4)-NH | 8.22 | 1.3 | 0 | Arg(8)-δH2 | 3.09 | 9.1 | 0.01 |
| Thr(4)-αH | 4.23 | 9.1 | 0.01 | Arg(8)-δH3 | 7.12 | 0 | 0 |
| Thr(4)-βH2 | 4.09 | | | Asp(9)-NH | 8.40 | 5.1 | 0.01 |
| Thr(4)-γH | 1.08 | 8.4 | 0.01 | Asp(9)-αH | 4.56 | 10.8 | 0.01 |
| Asn(5)-NH | 8.48 | 3.2 | 0 | Asp(9)-βH2 | 2.51 | 7.4 | 0.01 |
| Asn(5)-αH | 4.69 | | | Asp(9)-βH3 | 2.66 | 7.4 | 0.01 |
| Asn(5)-βH2 | 2.77 | 8.6 | 0.01 | Gly(10)-NH | 7.82 | 6.2 | 0.01 |
| Asn(5)-βH3 | 2.68 | 8.6 | 0.01 | Gly(10)-αH2 | 3.63 | 8.6 | 0.01 |
| | | | | Gly(10)-αH3 | 3.70 | 5.1 | 0.01 |

**Table A4-23. Differences in chemical shifts of Ligand 7 observed upon mixing the peptide Target 20 and Ligand 7**

| Ligand 1 | $\delta_H$ | Δδ [Hz] | Δδ [ppm] | | $\delta_H$ | Δδ [Hz] | Δδ [ppm] |
|---|---|---|---|---|---|---|---|
| Glu(1)-NH | - | | | Val(5)-NH | 7.99 | 25.6 | 0.03 |
| Glu(1)-αH | 3.85 | 3.1 | 0 | Val(5)-αH | 3.91 | 15.1 | 0.02 |
| Glu(1)-βH2 | 1.88 | | | Val(5)-βH | 1.93 | | 0 |
| Glu(1)-γH | 2.19 | | | Val(5)-γH2 | 0.81 | 7.3 | 0.01 |
| Asp(2)-NH | 8.67 | 2.6 | 0 | Glu(6)-NH | 8.32 | 33.4 | 0.04 |
| Asp(2)-αH | 4.53 | 4.7 | 0.01 | Glu(6)-αH | 4.24 | 11.0 | 0.01 |
| Asp(2)-βH2 | 2.59 | 15.1 | 0.02 | Glu(6)-βH2 | 1.84 | | |
| Asp(2)-βH3 | 2.46 | 17.2 | 0.02 | Glu(6)-βH3 | 1.97 | | |
| Tyr(3)-NH | 8.17 | 21.9 | 0.03 | Glu(6)-γH | 2.27 | | |
| Tyr(3)-αH | 4.43 | 17.2 | 0.02 | Glu(7)-NH | 8.32 | 24.0 | 0.03 |
| Tyr(3)-βH2 | 2.94 | 9.9 | 0.01 | Glu(7)-αH | 4.24 | 11.0 | 0.01 |
| Tyr(3)-βH3 | 2.81 | 3.1 | 0 | Glu(7)-βH2 | 1.84 | | |
| Tyr(3)-γH | 6.99 | 5.2 | 0.01 | Glu(7)-βH3 | 1.97 | | |
| Tyr(3)-γH | 6.69 | 6.3 | 0.01 | Glu(7)- γH | 2.27 | | |
| Glu(4)-NH | 8.14 | 1.6 | 0 | Gly(8)-NH | 8.00 | 3.1 | 0 |
| Glu(4)-αH | 4.18 | 8.4 | 0.01 | Gly(8)-αH | 3.65 | 6.8 | 0.01 |
| Glu(4)-βH2 | 1.81 | | | | | | |
| Glu(4)-βH3 | 1.92 | | | | | | |
| Glu(4)-γH | 2.19 | | | | | | |

## Ligand titrations with target/dilutions

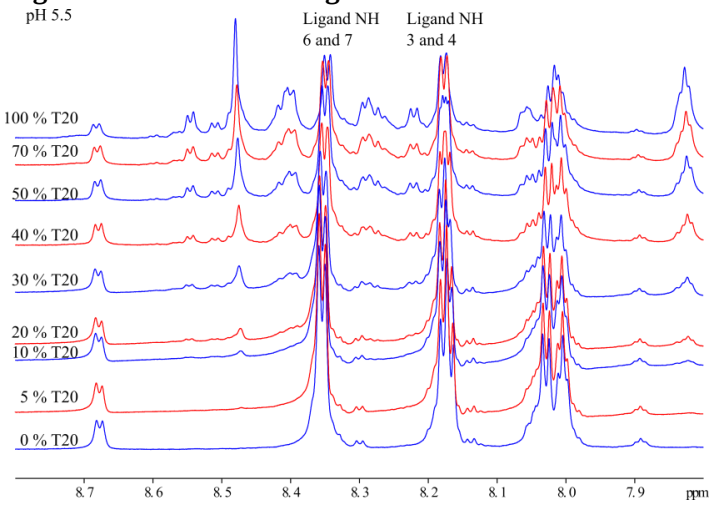



**Figure A4.6. Titration of Ligand 7 with Target 20, amide region at pH 5.5.**

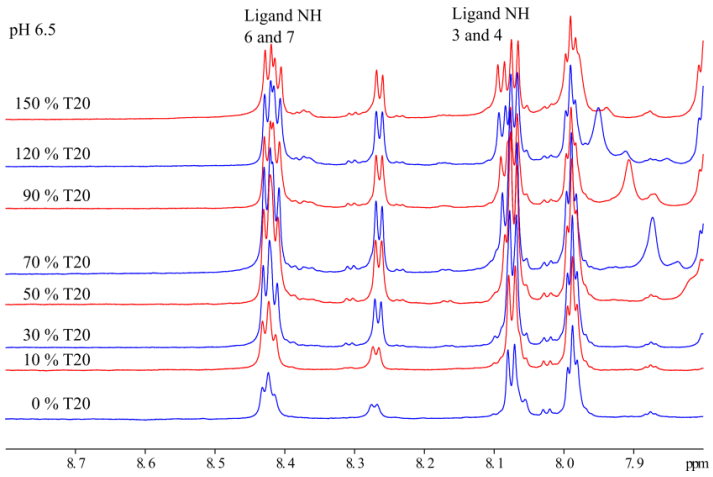



**Figure A4.7. Titration of Ligand 7 with Target 20, amide region at pH 6.5.**

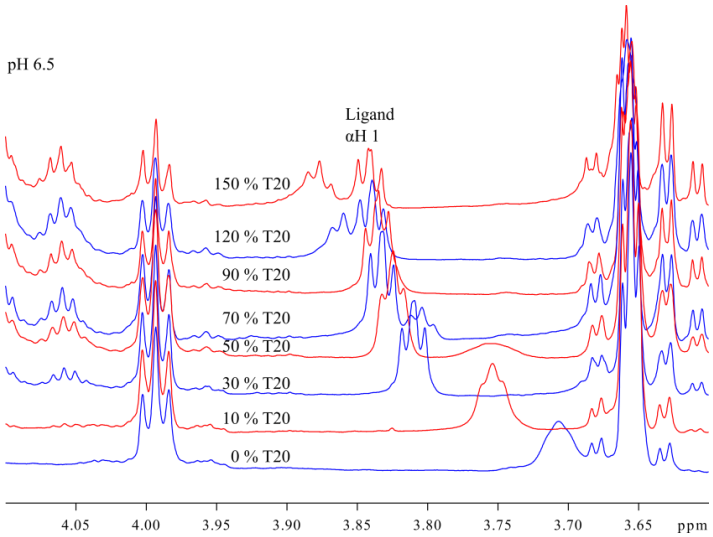



**Figure A4.8. Titration of Ligand 7 with Target 20, alpha region at pH 6.5.**

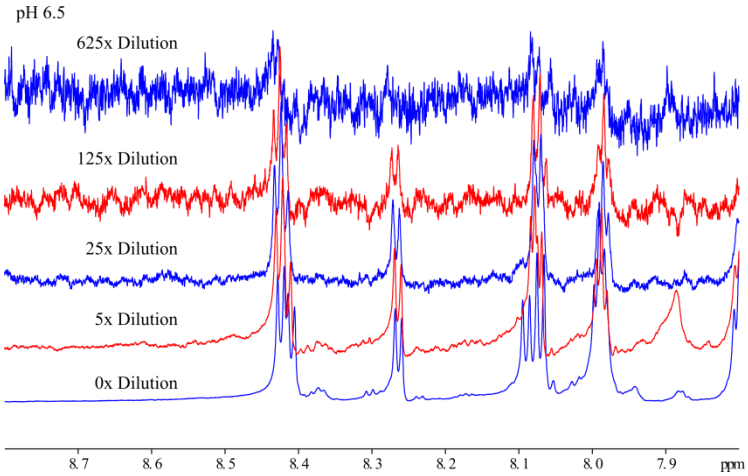



**Figure A4.9. Dilution of mixture of Ligand 7 and Target 20, amide region at pH 6.5.**

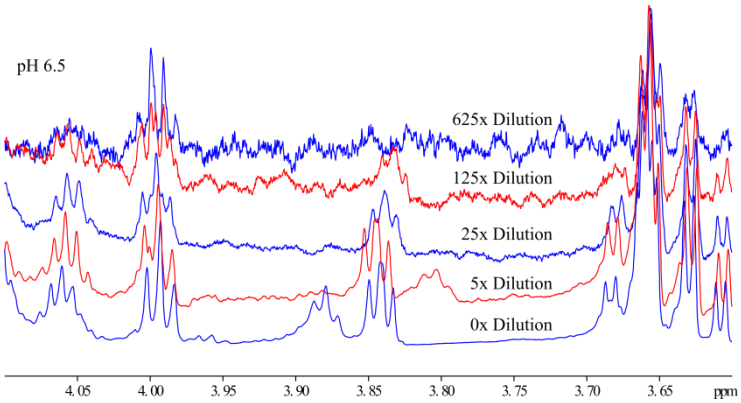


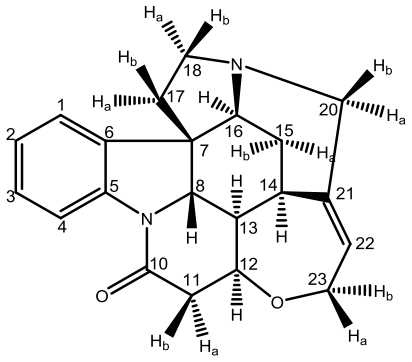

**Figure A4.10. Dilution of mixture of Ligand 7 and Target 20, alpha region at pH 6.5.**

## A4.3. Other compounds

### A4.3.1. Strychnine

**Table A4-24. NMR assignment of strychnine**

| # | $\delta_H$ [ppm] (int,mult,*J* [Hz]) | $\delta_C$ [ppm] |
|---|---|---|
| 1 | 7.12, 1H, d, 7.6 | 122.3 |
| 2 | 7.05, 1H, t, 7.6, | 124.3 |
| 3 | 7.21, 1H, t, 7.8 | 128.6 |
| 4 | 8.05, 1H, d, 8.0 | 116.2 |
| 5 | - | 142.3 |
| 6 | - | 132.9 |
| 7 | - | 51.9 |
| 8 | 3.82, 1H, d, 10.4 | 60.1 |
| 9-N | - | - |
| 10 | - | 169.4 |
| 11a | 3.09, 1H, m | 42.4 |
| 11b | 2.63, 1H, dd, 17.5, 3.4 | 42.4 |
| 12 | 4.25, 1H, dt, 8.2, 3.4 | 77.5 |
| 13 | 1.23, 1H, dt, 10.5, 3.2 | 48.1 |
| 14 | 3.10, 1H, m | 31.5 |
| 15a | 2.31, 1H, dt, 14.4, 4.3 | 26.7 |
| 15b | 1.41, 1H, br. d, 14.5 | 26.7 |
| 16 | 3.90, 1H, br. s | 60.1 |
| 17a/b | 1.84, 2H, m | 42.8 |
| 18a | 3.15, 1H, m | 50.2 |
| 18b | 2.83, 1H, m | 50.2 |
| 19-N | - | - |
| 20a | 3.66, 1H, d, 14.9 | 52.6 |
| 20b | 2.69, 1H, d, 14.6 | 52.6 |
| 21 | - | 140.6 |
| 22 | 5.86, 1H, m | 127.3 |
| 23a | 4.10, 1H, dd, 13.7, 7.1 | 64.6 |
| 23b | 4.02, 1H, dd, 13.7, 5.9 | 64.6 |



## NOE

**Table A4-25. $\tau_c$ determination. PANIC NOE and ROE intensities for strychnine ($\tau_m$ = 300 ms). Corr. factor given by equation 3.14 in the thesis. Corr. rel is the correction by including $\tau_c$ compared to the normal PANIC approach.**

| H1 | H2 | PANIC | | Corr. factor | ROE | NOE/ROE | $\tau_c$ [ps] | Corr. Rel | Rel. NOE | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | NOE | T-ROE | | | | | | NOE | NOE/ROE |
| 4 | 3 | 13.6 | 13.9 | 1.00 | 14.15 | 0.96 | 31.09 | 0.99 | 1.45 | 1.44 |
| 15a | 13 | 30.2 | 31.3 | 0.99 | 32.21 | 0.94 | 39.83 | 1.03 | 1.27 | 1.31 |
| 15a | 15b | 128.8 | 131.8 | 1.00 | 134.81 | 0.96 | 33.52 | 1.00 | 1.00 | 1.00 |
| 15a | 14 | 14.4 | 14.8 | 1.01 | 15.19 | 0.95 | 35.83 | 1.01 | 1.44 | 1.45 |
| 15a | 16 | 17.4 | 17.7 | 1.02 | 18.43 | 0.94 | 37.46 | 1.02 | 1.40 | 1.42 |
| 13 | 15a | 24.6 | 25.2 | 1.00 | 25.63 | 0.96 | 31.25 | 0.99 | 1.32 | 1.30 |
| 13 | 14 | 20.4 | 21.7 | 1.01 | 23.03 | 0.89 | 55.05 | 1.07 | 1.36 | 1.46 |
| 13 | 8 | 3.7 | 3.7 | 1.02 | 3.91 | 0.93 | 41.54 | 1.03 | 1.81 | 1.87 |
| 13 | 12 | 22.5 | 23.1 | 1.03 | 24.44 | 0.92 | 45.64 | 1.05 | 1.34 | 1.40 |
| 18b | 17b | 15.2 | 15.5 | 1.00 | 15.75 | 0.96 | 29.90 | 0.98 | 1.43 | 1.40 |
| 18b | 18a | 100.9 | 102.9 | 1.00 | 104.73 | 0.96 | 30.07 | 0.98 | 1.04 | 1.02 |
| 18b | 8 | 33.7 | 34.6 | 1.00 | 35.64 | 0.95 | 36.89 | 1.01 | 1.25 | 1.27 |
| 18b | 22 | 5.2 | 5.4 | 1.04 | 5.84 | 0.89 | 54.45 | 1.07 | 1.71 | 1.83 |
| 20a | 20b | 109.6 | 108.0 | 1.00 | 106.92 | 1.02 | - | - | 1.03 | - |
| 20a | 15b | 32.5 | 31.8 | 1.01 | 31.41 | 1.03 | - | - | 1.26 | - |
| 16 | 1 | 23.4 | 21.1 | 1.05 | 19.63 | 1.19 | - | - | 1.33 | - |
| 16 | 15b | 13.2 | 12.9 | 1.01 | 12.70 | 1.04 | - | - | 1.46 | - |
| 16 | 17a | 9.6 | 9.3 | 1.02 | 9.09 | 1.06 | - | - | 1.54 | - |
| 16 | 15a | 13.7 | 12.8 | 1.03 | 12.31 | 1.11 | - | - | 1.45 | - |
| 12 | 23a | 22.6 | 22.4 | 1.00 | 22.34 | 1.01 | - | - | 1.34 | - |
| 12 | 11a | 29.9 | 29.0 | 1.01 | 28.29 | 1.06 | - | - | 1.28 | - |
| 12 | 11b | 3.8 | 3.8 | 1.01 | 3.79 | 1.00 | - | - | 1.80 | - |
| 12 | 13 | 22.3 | 20.6 | 1.04 | 19.83 | 1.12 | - | - | 1.34 | - |
| 22 | 23b | 14.5 | 14.1 | 1.01 | 13.80 | 1.05 | - | - | 1.44 | - |
| 22 | 20b | 18.3 | 17.7 | 1.04 | 17.73 | 1.03 | - | - | 1.38 | - |
| 8 | 18b | 25.5 | 25.1 | 1.00 | 24.75 | 1.03 | - | - | 1.31 | - |
| 8 | 17a | 17.8 | 16.9 | 1.02 | 16.27 | 1.09 | - | - | 1.39 | - |

## RDC

**Table A4-26. RDC extraction of strychnine (Polymer 6.2 in the thesis). $^2$H splitting = 112.3 Hz ($CDCl_3$)**

| Polymer 6.2 | | | | |
|---|---|---|---|---|
| Nuc 1 | Nuc 2 | *J* [Hz] | *T* [Hz] | *D* [Hz] |
| C1 | H1 | 158.3 | 184.9 | 26.6 |
| C2 | H2 | 160.7 | 174.5 | 13.8 |
| C3 | H3 | 159.8 | 183.6 | 23.8 |
| C4 | H4 | 168.5 | 196.8 | 28.3 |
| C8 | H8 | 144.9 | 130.6 | -14.4 |
| C11 | H11a | 135.5 | 147.9 | 12.3 |
| C11 | H11b | 125.7 | 99.9 | -25.8 |
| C12 | H12 | 149.3 | 120.9 | -28.4 |
| C13 | H13 | 125.0 | 120.7 | -4.3 |
| C14 | H14 | 131.6 | 103.7 | -27.9 |
| C15 | H15a | 131.6 | 138.6 | 7 |
| C15 | H15b | 130.4 | 137.8 | 7.5 |
| C16 | H16 | 146.5 | 161.7 | 15.2 |
| C18 | H18a | 146.4 | 147.9 | 1.6 |
| C18 | H18b | 131.2 | 145.9 | 14.7 |
| C20 | H20a | 138.5 | 122.5 | -16 |
| C20 | H20b | 138.2 | 134.1 | -4.1 |
| C22 | H22 | 159.4 | 160.1 | 0.7 |
| H20a | H20b | -14.8 | -51.7 | -36.9 |
| H11a | H11b | -17.2 | -39.3 | -22.1 |
| H3 | H4 | 8.4 | 12.8 | 4.4 |
| H3 | H2 | 7.4 | 11.0 | 3.6 |
| H3 | H1 | 1.5 | 3.1 | 1.7 |
| H22 | H23a | 6.8 | 15.5 | 8.8 |
| H22 | H20b | -0.7 | 5.8 | 6.5 |
| H2 | H4 | 1.0 | 1.2 | 0.2 |
| H1 | H3 | 1.3 | 2.5 | 1.2 |
| H1 | H2 | 7.1 | 11.9 | 4.8 |
| H4 | H3 | 7.5 | 13.1 | 5.7 |
| H4 | H2 | 1.1 | 2.8 | 1.8 |
| H12 | H23a | -0.3 | -1.0 | -0.7 |
| H12 | H11b | 3.3 | -1.1 | -4.5 |
| H16 | H20b | -0.3 | 0.1 | 0.4 |
| H16 | H15a | 3.9 | -3.0 | -6.9 |
| H16 | H15b | 2.2 | -0.1 | -2.3 |
| H8 | H13 | 10.5 | 7.6 | -3 |
| H8 | H12 | -0.1 | -1.8 | -1.7 |
| H16 | H18a | -0.3 | -4.2 | -4 |
| H13 | H14 | -0.4 | 2.3 | 2.7 |
| H13 | H15a | 0.4 | 2.0 | 1.6 |
| H13 | H15b | -0.6 | 4.5 | 5.1 |
| H14 | H22 | -3.0 | -3.4 | -0.3 |
| H14 | H20b | -0.5 | -0.7 | -0.2 |
| H14 | H15b | 1.9 | 4.6 | 2.7 |

**Table A4-27. RDC extraction of strychnine (Polymer 8.1 in the thesis). $^2$H splitting = 103.4 Hz ($CDCl_3$)**

| Polymer 8.1 | | | | |
|---|---|---|---|---|
| C | H | *J* [Hz] | *T* [Hz] | *D* [Hz] |
| C1 | H1 | 158.3 | 177.3 | 19 |
| C2 | H2 | 160.7 | 174.7 | 14 |
| C3 | H3 | 159.8 | 178.7 | 18.8 |
| C4 | H4 | 168.5 | 187.2 | 18.7 |
| C8 | H8 | 144.9 | 134.0 | -11 |
| C11 | H11a | 135.5 | 140.6 | 5.1 |
| C11 | H11b | 125.7 | 104.1 | -21.7 |
| C12 | H12 | 149.3 | 124.1 | -25.2 |
| C13 | H13 | 125.0 | 121.2 | -3.8 |
| C14 | H14 | 131.6 | 106.5 | -25.1 |
| C15 | H15a | 131.6 | 134.6 | 3.1 |
| C15 | H15b | 130.4 | 136.5 | 6.1 |
| C16 | H16 | 146.5 | 161.4 | 15 |
| C18 | H18a | 146.4 | 143.5 | -2.9 |
| C18 | H18b | 131.2 | 144.4 | 13.2 |
| C20 | H20a | 138.5 | 124.1 | -14.4 |
| C20 | H20b | 138.2 | 134.7 | -3.5 |
| C22 | H22 | 159.4 | 159.4 | 0.1 |
| H1 | H2 | 7.2 | 11.1 | 3.9 |
| H1 | H3 | 1.3 | 2.9 | 1.6 |
| H2 | H3 | 7.4 | 10.8 | 3.4 |
| H2 | H4 | 1.1 | 1.4 | 0.3 |
| H3 | H4 | 8.3 | 12.9 | 4.6 |
| H8 | H12 | 0.1 | -3.2 | -3.3 |
| H8 | H13 | 10.5 | 6.3 | -4.2 |
| H8 | H16 | -0.1 | 2.0 | 2.1 |
| H12 | H11b | 3.3 | 0.0 | -3.2 |
| H12 | H23a | -0.2 | 1.2 | 1.5 |
| H12 | H23b | -0.4 | 9.1 | 9.4 |
| H13 | H14 | -0.5 | 2.3 | 2.8 |
| H13 | H15a | 0.6 | 2.7 | 2.1 |
| H13 | H15b | -0.6 | 7.5 | 8.1 |
| H14 | H15b | 1.8 | 5.6 | 3.8 |
| H14 | H16 | 0.7 | 1.8 | 1.1 |
| H14 | H20b | -0.6 | -1.0 | -0.4 |
| H14 | H22 | -3.1 | -3.7 | -0.7 |
| H16 | H15a | 3.9 | 0.9 | -3 |
| H16 | H15b | 2.2 | -0.3 | -2.5 |
| H16 | H17a/H17b | -0.1 | -4.2 | -4.1 |
| H16 | H18a | -0.2 | -1.9 | -1.7 |
| H16 | H20b | -0.3 | 1.6 | 1.9 |
| H22 | H20b | -0.7 | 4.7 | 5.4 |
| H22 | H23a | 7.0 | 12.3 | 5.3 |
| H22 | H23b | 6.0 | 3.9 | -2.1 |

**Table A4-28. RDCs in Hz back-calculated from different assignments of strychnine using SVD (Polymer 6.2 in the thesis).**

| C | H | Exp. | Right | Dia-11 | Dia-15 | Dia-18 | Dia-20 | Dia-23 |
|---|---|---|---|---|---|---|---|---|
| C1 | H1 | 19.0 | 18.0 | 11.4 | 18.4 | 19.8 | 18.8 | 18.0 |
| C2 | H2 | 14.0 | 14.8 | 13.6 | 14.4 | 9.1 | 13.9 | 14.8 |
| C3 | H3 | 18.8 | 18.2 | 15.5 | 17.8 | 18.1 | 15.8 | 18.2 |
| C4 | H4 | 18.7 | 18.3 | 11.7 | 18.7 | 20.0 | 19.2 | 18.3 |
| C8 | H8 | -11.0 | -10.3 | -0.8 | -10.6 | -12.1 | -10.5 | -10.3 |
| C11 | H11b | 5.1 | 7.5 | -12.9 | 7.2 | 11.5 | 6.0 | 7.5 |
| C11 | H11a | -21.7 | -23.7 | -2.3 | -23.8 | -23.4 | -22.8 | -23.7 |
| C12 | H12 | -25.2 | -25.5 | -25.7 | -25.7 | -21.8 | -25.8 | -25.5 |
| C13 | H13 | -3.8 | -3.6 | 0.4 | -4.3 | -2.5 | -5.8 | -3.6 |
| C14 | H14 | -25.1 | -24.2 | -24.3 | -24.2 | -21.6 | -24.0 | -24.2 |
| C15 | H15a | 3.1 | 3.2 | 5.7 | 5.2 | 3.1 | 6.7 | 3.2 |
| C15 | H15b | 6.1 | 5.9 | 6.2 | 4.3 | 7.7 | 3.0 | 5.9 |
| C16 | H16 | 15.0 | 14.1 | 12.7 | 13.9 | 9.0 | 13.8 | 14.1 |
| C18 | H18a | -2.9 | -4.3 | -10.7 | -4.8 | 8.4 | -6.5 | -4.3 |
| C18 | H18b | 13.2 | 13.6 | 12.9 | 13.4 | 1.5 | 13.5 | 13.6 |
| C20 | H20a | -14.4 | -12.5 | -3.7 | -11.5 | -13.8 | -5.5 | -12.5 |
| C20 | H20b | -3.5 | -3.1 | 0.1 | -3.8 | -1.4 | -8.1 | -3.1 |
| C22 | H22 | 0.1 | 2.2 | 2.5 | 1.4 | 4.9 | -0.7 | 2.2 |
| | | Q-factor | 0.079 | 0.547 | 0.096 | 0.334 | 0.201 | 0.079 |

**Table A4-29. RDCs in Hz back-calculated from different assignments of strychnine using SVD (Polymer 6.2 in the thesis).**

| Nuc 1 | Nuc 2 | Exp. | Right | Dia-11 | Dia-15 | Dia-18 | Dia-20 | Dia-23 |
|---|---|---|---|---|---|---|---|---|
| C1 | H1 | 19.0 | 15.8 | 13.4 | 19.2 | 16.8 | 16.1 | 14.7 |
| C2 | H2 | 14.0 | 14.5 | 15.5 | 12.1 | 11.8 | 15.8 | 14.4 |
| C3 | H3 | 18.8 | 17.9 | 17.4 | 18.4 | 18.5 | 14.8 | 20.3 |
| C4 | H4 | 18.7 | 16.0 | 13.6 | 19.5 | 17.0 | 16.5 | 14.9 |
| C8 | H8 | -11.0 | -7.9 | -3.6 | -9.7 | -10.4 | -6.4 | -4.0 |
| C11 | H11b | 5.1 | 4.6 | -17.1 | 5.4 | 7.1 | 2.1 | 6.2 |
| C11 | H11a | -21.7 | -21.1 | 1.5 | -22.3 | -22.6 | -19.2 | -18.3 |
| C12 | H12 | -25.2 | -25.5 | -26.3 | -27.1 | -23.3 | -27.0 | -25.4 |
| C13 | H13 | -3.8 | -2.5 | -0.1 | -3.8 | -2.9 | -4.3 | 3.5 |
| C14 | H14 | -25.1 | -24.1 | -25.1 | -25.7 | -22.2 | -25.1 | -25.6 |
| C15 | H15a | 3.1 | 3.6 | 3.1 | 5.7 | 3.5 | 6.7 | -1.0 |
| C15 | H15b | 6.1 | 6.2 | 7.2 | 7.5 | 6.6 | 3.0 | 12.1 |
| C16 | H16 | 15.0 | 13.7 | 14.5 | 12.0 | 11.1 | 15.5 | 13.1 |
| C18 | H18a | -2.9 | -6.2 | -8.2 | -6.1 | 10.5 | -9.9 | -2.3 |
| C18 | H18b | 13.2 | 13.3 | 14.4 | 11.7 | -3.2 | 15.4 | 12.7 |
| C20 | H20a | -14.4 | -10.5 | -8.7 | -8.2 | -12.2 | -4.5 | -14.9 |
| C20 | H20b | -3.5 | -2.2 | -0.1 | -3.3 | -2.2 | -5.4 | 4.0 |
| C22 | H22 | 0.1 | 2.4 | 3.4 | 1.9 | 3.2 | -1.0 | 8.7 |
| H1 | H3 | 1.5 | 1.1 | 1.0 | 1.4 | 1.3 | 1.0 | 1.2 |
| H1 | H3 | 1.6 | 1.1 | 1.0 | 1.4 | 1.3 | 1.0 | 1.2 |
| H1 | H2 | 3.9 | 6.0 | 5.8 | 6.1 | 6.2 | 4.9 | 6.8 |
| H2 | H3 | 3.1 | 5.3 | 4.5 | 6.5 | 5.6 | 5.5 | 4.9 |
| H2 | H4 | 0.2 | 0.9 | 0.9 | 0.9 | 0.8 | 1.0 | 0.8 |
| H2 | H3 | 3.6 | 5.3 | 4.5 | 6.5 | 5.6 | 5.5 | 4.9 |
| H2 | H4 | 0.3 | 0.9 | 0.9 | 0.9 | 0.8 | 1.0 | 0.8 |
| H3 | H4 | 4.4 | 4.8 | 5.1 | 4.0 | 3.9 | 5.2 | 4.7 |
| H3 | H4 | 4.8 | 4.8 | 5.1 | 4.0 | 3.9 | 5.2 | 4.7 |
| H8 | H12 | -3.3 | -2.1 | -2.3 | -2.3 | -1.8 | -2.3 | -2.0 |
| H8 | H16 | 2.1 | 0.6 | 0.7 | 0.7 | 0.4 | 0.9 | 0.5 |
| H8 | H13 | -3.8 | -5.1 | -4.6 | -5.4 | -5.1 | -5.1 | -4.5 |
| H8 | H13 | -4.6 | -5.1 | -4.6 | -5.4 | -5.1 | -5.1 | -4.5 |
| H11a | H12 | -3.2 | -5.2 | 0.8 | -5.1 | -5.3 | -4.6 | -6.0 |
| H12 | H23a | 1.5 | 1.1 | 1.3 | 1.0 | 1.1 | 0.8 | 8.3 |
| H12 | H23b | 9.4 | 7.6 | 7.1 | 8.1 | 8.0 | 6.3 | 1.8 |
| H13 | H14 | 2.8 | 4.2 | 3.6 | 5.0 | 4.6 | 3.9 | 3.3 |
| H13 | H15a | 2.1 | 1.7 | 1.5 | 6.5 | 1.8 | 1.9 | 1.5 |
| H13 | H15b | 8.1 | 5.5 | 5.4 | 2.2 | 4.8 | 7.0 | 4.3 |
| H14 | H15b | 3.8 | 4.4 | 4.4 | 4.3 | 3.9 | 4.2 | 4.3 |
| H14 | H16 | 1.1 | 0.1 | -0.1 | -0.1 | 0.1 | 0.0 | 0.0 |
| H14 | H20b | -0.4 | -0.7 | -0.7 | -0.7 | -0.5 | 0.3 | -0.3 |
| H14 | H22 | -0.7 | -0.7 | -0.4 | -0.8 | -0.8 | -0.8 | -0.2 |
| H15a | H16 | -3.0 | -4.4 | -4.9 | -4.7 | -3.9 | -4.7 | -5.5 |
| H15b | H16 | -2.5 | -4.3 | -5.1 | -5.0 | -3.3 | -5.3 | -3.7 |
| H16 | H17a/H17b | -4.1 | -5.0 | -5.4 | -6.2 | -6.1 | -5.9 | -5.0 |
| H16 | H18a | -1.7 | -1.6 | -1.2 | -1.7 | -0.4 | -1.5 | -1.1 |
| H16 | H20b | 1.9 | 1.1 | 1.2 | 0.9 | 1.0 | 1.7 | 1.3 |
| H20b | H22 | 5.4 | 4.9 | 4.5 | 4.6 | 4.5 | 1.6 | 4.5 |
| H22 | H23a | 5.3 | 4.1 | 3.0 | 4.6 | 4.7 | 3.7 | -0.7 |
| H22 | H23b | -2.1 | -1.5 | -1.8 | -1.5 | -1.0 | -2.3 | 4.1 |
| | | *Q*-factor | 0.162 | 0.555 | 0.214 | 0.365 | 0.272 | 0.366 |

## Chiral polymer (L-TEG-Phe/DMAA)

**Table A4-30. Output from MSPIN upon entering the extracted RDCs of strychnine aligned in the L-TEG-Phe/DMAA**

| 1 | Exp Hz | Comp Hz | 2 | Exp Hz | Comp Hz | 3 | Exp Hz | Comp Hz |
|---|---|---|---|---|---|---|---|---|
| C22,H45 | 0.9 | -1.2 | C22,H45 | 2.2 | -1.1 | C22,H45 | 2.1 | -0.94 |
| C2,H27 | 8.2 | 9.5 | C2,H27 | 8.5 | 10.1 | C2,H27 | 7.8 | 9.94 |
| C1,H26 | -21.1 | -19.6 | C1,H26 | -22.6 | -20.5 | C1,H26 | -25.1 | -22.68 |
| C4,H29 | -22.9 | -19.9 | C4,H29 | -22.7 | -20.8 | C4,H29 | -26.4 | -23.07 |
| C12,H33 | 10.4 | 13.8 | C12,H33 | 10.5 | 14.5 | C12,H33 | 10.9 | 15.35 |
| C16,H38 | 7.5 | 7.7 | C16,H38 | 7.6 | 8.1 | C16,H38 | 7.9 | 7.36 |
| C8,H30 | 9.7 | 14.8 | C8,H30 | 11.3 | 14.7 | C8,H30 | 10.7 | 14.29 |
| C20,H43 | 1.7 | -2.1 | C20,H43 | -0.4 | -3.2 | C20,H43 | -1.4 | -4.21 |
| C20,H44 | 8.2 | 4.9 | C20,H44 | 7.6 | 5.0 | C20,H44 | 5.3 | 5.09 |
| C18,H41 | 5.6 | -1.0 | C18,H41 | 3.9 | -0.8 | C18,H41 | 5.9 | -1.45 |
| C18,H42 | 13.0 | 7.7 | C18,H42 | 12.8 | 8.0 | C18,H42 | 11.3 | 7.14 |
| C13,H34 | 2.9 | 6.4 | C13,H34 | 1.7 | 6.5 | C13,H34 | 2.4 | 6.62 |
| C11,H32 | -9.1 | -6.3 | C11,H32 | -7.5 | -6.2 | C11,H32 | -9.9 | -7.58 |
| C11,H31 | 17.3 | 16.7 | C11,H31 | 17.3 | 16.4 | C11,H31 | 17.3 | 16.01 |
| C14,H35 | 13.5 | 13.0 | C14,H35 | 15.3 | 13.7 | C14,H35 | 15 | 15.1 |
| *Q*-factor | | 0.284 | *Q*-factor | | 0.247 | *Q*-factor | | 0.259 |
| Alignment tensor | | | Alignment tensor | | | Alignment tensor | | |
| A'x=-1.263e-04 | | | A'x=-1.421e-04 | | | A'x=-1.544e-04 | | |
| A'y=-2.468e-04 | | | A'y=-2.444e-04 | | | A'y=-2.456e-04 | | |
| A'z= 3.732e-04 | | | A'z= 3.865e-04 | | | A'z= 4.001e-04 | | |
| Saupe tensor | | | Saupe tensor | | | Saupe tensor | | |
| S'x=-1.895e-04 | | | S'x=-2.131e-04 | | | S'x=-2.317e-04 | | |
| S'y=-3.703e-04 | | | S'y=-3.666e-04 | | | S'y=-3.684e-04 | | |
| S'z= 5.597e-04 | | | S'z= 5.797e-04 | | | S'z= 6.001e-04 | | |
| Alignment tensor eigenvectors | | | Alignment tensor eigenvectors | | | Alignment tensor eigenvectors | | |
| e[x]=( 0.765,-0.639, 0.081) | | | e[x]=( 0.780,-0.626, 0.021) | | | e[x]=(-0.820, 0.561, 0.115) | | |
| e[y]=( 0.143, 0.046,-0.989) | | | e[y]=( 0.103, 0.095,-0.990) | | | e[y]=( 0.015,-0.180, 0.983) | | |
| e[z]=( 0.628, 0.768, 0.127) | | | e[z]=( 0.618, 0.774, 0.138) | | | e[z]=( 0.572, 0.808, 0.139) | | |
| Alignment tensor in laboratory coordinates: | | | Alignment tensor in laboratory coordinates: | | | Alignment tensor in laboratory coordinates: | | |
| [ 6.845e-05,2.401e-04,5.679e-05] | | | [ 5.839e-05,2.517e-04,5.550e-05] | | | [ 2.714e-05,2.567e-04,4.290e-05] | | |
| [ 2.401e-04,1.677e-04,5.403e-05] | | | [ 2.517e-04,1.739e-04,6.611e-05] | | | [ 2.567e-04,2.047e-04,7.866e-05] | | |
| [ 5.679e-05,5.403e-05,-2.361e-04] | | | [ 5.550e-05,6.611e-05,-2.323e-04] | | | [ 4.290e-05,7.866e-05,-2.319e-04] | | |
| SVD condition number is 1.742e+00 | | | SVD condition number is 1.742e+00 | | | SVD condition number is 1.742e+00 | | |
| Axial component Aa = 5.597e-04 | | | Axial component Aa = 5.797e-04 | | | Axial component Aa = 6.001e-04 | | |
| Rhombic component Ar = 1.205e-04 | | | Rhombic component Ar = 1.024e-04 | | | Rhombic component Ar = 9.119e-05 | | |
| rhombicity R = 2.153e-01 | | | rhombicity R = 1.766e-01 | | | rhombicity R = 1.520e-01 | | |
| Asimmetry parameter etha =3.230e-01 | | | Asimmetry parameter etha =2.648e-01 | | | Asimmetry parameter etha =2.279e-01 | | |
| GDO = 6.630e-04 | | | GDO = 6.810e-04 | | | GDO = 7.019e-04 | | |
| Euler Angles (degrees) | | | Euler Angles (degrees) | | | Euler Angles (degrees) | | |
| Set 1 | | | Set 1 | | | Set 1 | | |
| (80.6,-38.9,10.6) | | | (79.9,-38.1,7.5) | | | (80.2,-34.9,178.9) | | |
| Set 2 | | | Set 2 | | | Set 2 | | |
| (-99.4,218.9,-169.4) | | | (-100.1,218.1,-172.5) | | | (-99.8,214.9,-1.1) | | |

## Chiral polymer (D-TEG-Phe/DMAA)

**Table A4-31. Output from MSPIN upon entering the extracted RDCs of strychnine aligned in the D-TEG-Phe/DMAA**

| **4** | **Exp Hz** | **Comp Hz** | **5** | **Exp Hz** | **Comp Hz** | **6** | **Exp Hz** | **Comp Hz** |
|---|---|---|---|---|---|---|---|---|
| C22,H45 | 3.3 | 1.0 | C22,H45 | 3.1 | 0.4 | C22,H45 | 3.5 | 0.45 |
| C2,H27 | 6.0 | 6.5 | C2,H27 | 6.3 | 6.7 | C2,H27 | 7.6 | 8.35 |
| C1,H26 | -13.7 | -12.0 | C1,H26 | -14.2 | -14.2 | C1,H26 | -15.5 | -14.47 |
| C4,H29 | -12.5 | -12.2 | C4,H29 | -15.3 | -14.5 | C4,H29 | -16.9 | -14.73 |
| C12,H33 | 6.9 | 10.5 | C12,H33 | 8.4 | 12.1 | C12,H33 | 6.8 | 11.11 |
| C16,H38 | 5.8 | 6.0 | C16,H38 | 6.1 | 5.9 | C16,H38 | 7.3 | 7.27 |
| C8,H30 | 9.5 | 13.4 | C8,H30 | 11.1 | 13.4 | C8,H30 | 10.2 | 13.27 |
| C20,H43 | 0.8 | -0.9 | C20,H43 | -0.5 | -1.4 | C20,H43 | -1.6 | -2.65 |
| C20,H44 | 10.1 | 5.7 | C20,H44 | 10.4 | 5.5 | C20,H44 | 8.9 | 5.4 |
| C18,H41 | 7.4 | 2.4 | C18,H41 | 7.1 | 2.0 | C18,H41 | 8.2 | 1.36 |
| C18,H42 | 10.0 | 6.2 | C18,H42 | 9.1 | 6.0 | C18,H42 | 10.6 | 7.3 |
| C13,H34 | 0.8 | 6.8 | C13,H34 | -1.2 | 6.6 | C13,H34 | -2 | 6.57 |
| C11,H32 | -4.9 | -1.8 | C11,H32 | -6.9 | -2.8 | C11,H32 | -6.8 | -2.9 |
| C11,H31 | 17.3 | 15.3 | C11,H31 | 17.3 | 15.5 | C11,H31 | 17.3 | 14.52 |
| C14,H35 | 9.9 | 8.3 | C14,H35 | 12.6 | 10.2 | C14,H35 | 10.1 | 9.49 |
| *Q*-factor | | 0.347 | *Q*-factor | | 0.345 | *Q*-factor | | 0.374 |
| Alignment tensor | | | Alignment tensor | | | Alignment tensor | | |
| A'x=-6.623e-05 | | | A'x=-7.895e-05 | | | A'x=-9.995e-05 | | |
| A'y=-2.233e-04 | | | A'y=-2.305e-04 | | | A'y=-2.125e-04 | | |
| A'z= 2.896e-04 | | | A'z= 3.095e-04 | | | A'z= 3.124e-04 | | |
| Saupe tensor | | | Saupe tensor | | | Saupe tensor | | |
| S'x=-9.935e-05 | | | S'x=-1.184e-04 | | | S'x=-1.499e-04 | | |
| S'y=-3.350e-04 | | | S'y=-3.457e-04 | | | S'y=-3.187e-04 | | |
| S'z= 4.344e-04 | | | S'z= 4.642e-04 | | | S'z= 4.686e-04 | | |
| Alignment tensor eigenvectors | | | Alignment tensor eigenvectors | | | Alignment tensor eigenvectors | | |
| e[x]=(-0.633, 0.761,-0.143) | | | e[x]=(-0.696, 0.714,-0.072) | | | e[x]=( 0.692,-0.705, 0.159) | | |
| e[y]=(-0.231,-0.009, 0.973) | | | e[y]=(-0.169,-0.065, 0.984) | | | e[y]=( 0.239, 0.016,-0.971) | | |
| e[z]=( 0.739, 0.649, 0.182) | | | e[z]=( 0.698, 0.697, 0.166) | | | e[z]=( 0.682, 0.709, 0.180) | | |
| Alignment tensor in laboratory coordinates: | | | Alignment tensor in laboratory coordinates: | | | Alignment tensor in laboratory coordinates: | | |
| [ 1.196e-04,1.702e-04,8.310e-05] | | | [ 1.060e-04,1.872e-04,7.006e-05] | | | [ 8.516e-05,1.989e-04,7.658e-05] | | |
| [ 1.702e-04,8.363e-05,4.343e-05] | | | [ 1.872e-04,1.089e-04,5.455e-05] | | | [ 1.989e-04,1.075e-04,5.424e-05] | | |
| [ 8.310e-05,4.343e-05,-2.032e-04] | | | [ 7.006e-05,5.455e-05,-2.149e-04] | | | [ 7.658e-05,5.424e-05,-1.927e-04] | | |
| SVD condition number is 1.742e+00 | | | SVD condition number is 1.742e+00 | | | SVD condition number is 1.742e+00 | | |
| Axial component Aa = 4.344e-04 | | | Axial component Aa = 4.642e-04 | | | Axial component Aa = 4.686e-04 | | |
| Rhombic component Ar = 1.571e-04 | | | Rhombic component Ar = 1.515e-04 | | | Rhombic component Ar = 1.125e-04 | | |
| rhombicity R = 3.617e-01 | | | rhombicity R = 3.265e-01 | | | rhombicity R = 2.401e-01 | | |
| Asimmetry parameter etha =5.426e-01 | | | Asimmetry parameter etha =4.897e-01 | | | Asimmetry parameter etha =3.601e-01 | | |
| GDO = 5.372e-04 | | | GDO = 5.672e-04 | | | GDO = 5.584e-04 | | |
| Euler Angles (degrees) | | | Euler Angles (degrees) | | | Euler Angles (degrees) | | |
| Set 1 | | | Set 1 | | | Set 1 | | |
| (74.4,-47.6,-159.9) | | | (76.6,-44.3,-166.4) | | | (75.8,-43.0,19.1) | | |
| Set 2 | | | Set 2 | | | Set 2 | | |
| (-105.6,227.6,20.1) | | | (-103.4,224.3,13.6) | | | (-104.2,223.0,-160.9) | | |

**Table A4-32. Comparison of the alignment frames obtained from MSPIN upon entering the extracted RDCs of strychnine aligned in the L- or D-TEG-Phe/DMAA**

| GCB | 1 | 2 | 3 | 4 | 5 | 6 | GCB | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 1.000 | 0.999 | 0.984 | 0.938 | 0.970 | 0.986 | 1 | 0 | 2.87 | 10.26 | 20.36 | 14.11 | 9.65 |
| **2** | | 1.000 | 0.992 | 0.920 | 0.957 | 0.977 | 2 | | 0 | 7.41 | 23.14 | 16.95 | 12.20 |
| **3** | | | 1.000 | 0.862 | 0.911 | 0.943 | 3 | | | 0 | 30.51 | 24.35 | 19.38 |
| **4** | | | | 1.000 | 0.994 | 0.980 | 4 | | | | 0 | 6.44 | 11.38 |
| **5** | | | | | 1.000 | 0.994 | 5 | | | | | 0 | 6.18 |
| **6** | | | | | | 1.000 | 6 | | | | | | 0 |

## A4.3.2. Quinine

**Table A4-33. NMR assignment of quinine**

| # | $\delta_H$ [ppm] (int,mult,*J* [Hz]) | $\delta_C$ [ppm] |
|---|---|---|
| 2' | 8.47 (1H,d,4.6) | 147.2 |
| 3' | 7.44 (1H,d,4.6) | 118.4 |
| 4' | | 148.0 |
| 5' | 7.21 (1H,m) | 121.3 |
| 6' | | 157.7 |
| 7' | 7.26 (1H,m) | 101.2 |
| 8' | 7.88 (1H,d,9.1) | 131.2 |
| 9' | | 143.4 |
| 10' | | 126.8 |
| 2a | 3.03 (1H,dd,13.7,10.1) | 56.9 |
| 2b | 2.61 (1H,m) | 56.9 |
| 3 | 2.23 (1H,br. t,10.0) | 39.8 |
| 4 | 1.77 (1H,m) | 27.8 |
| 5a | 1.70 (1H,m) | 21.5 |
| 5b | 1.45 (1H,m) | 21.5 |
| 6a | 3.42 (1H,dddd,12.9,10.6,5.0,2.2) | 43.1 |
| 6b | 2.60 (1H,m) | 43.1 |
| 7a | 1.72 (1H,m) | 27.6 |
| 7b | 1.49 (1H,m) | 27.6 |
| 8 | 3.07 (1H,m) | 59.9 |
| 9 | 5.51 (1H,d,3.2) | 71.6 |
| 10 | 5.74 (1H,ddd,17.3,10.1,7.5) | 141.7 |
| 11a | 4.96 (1H,m) | 114.3 |
| 11b | 4.92 (1H,m) | 114. |
| OMe | 3.9 (3H,s) | 55.6 |
| OH | 4.72 (1H,br. s) | - |

## NOE

**Table A4-34. Experimental distances in Å from relative NOE intensities compared to distances of the relevant structures for quinine ($\tau_m$ = 300 ms)**

| H1 | H2 | PANIC NOE | Rel. Distance | Exp. Distance | Theoretical distances | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | C1 | C2 | O3 | C7 | O8 | O4 | O10 |
| 2' | 3' | 14.43 | 1.39 | 2.47 | 2.49 | 2.50 | 2.50 | 2.49 | 2.47 | 2.47 | 2.48 |
| 9 | 3' | 5.89 | 1.62 | 2.86 | 2.28 | 3.68 | 3.65 | 3.55 | 3.62 | 3.62 | 3.61 |
| 9 | 5' | 39.25 | 1.18 | 2.09 | 3.79 | 2.09 | 2.13 | 2.34 | 2.29 | 2.29 | 2.07 |
| 9 | 6a | 13.02 | 1.42 | 2.51 | 2.13 | 2.16 | 2.67 | 2.23 | 2.69 | 2.69 | 3.97 |
| 9 | 8 | 11.35 | 1.45 | 2.57 | 3.04 | 3.04 | 2.56 | 3.05 | 2.53 | 2.53 | 2.24 |
| 9 | 7a | 7.36 | 1.56 | 2.76 | 2.60 | 2.71 | 3.63 | 2.44 | 3.65 | 3.65 | 2.85 |
| 6a | 9 | 15.87 | 1.37 | 2.43 | 2.13 | 2.16 | 2.67 | 2.23 | 2.69 | 2.69 | 3.97 |
| 6a | 6b | 105.52 | 1.00 | 1.77 | 1.76 | 1.76 | 1.77 | 1.76 | 1.77 | 1.77 | 1.75 |
| 6a | 5a | 23.46 | 1.28 | 2.28 | 2.35 | 2.34 | 2.33 | 2.37 | 2.33 | 2.33 | 2.33 |
| 6a | 5b | 3.85 | 1.74 | 3.07 | 2.97 | 2.98 | 2.92 | 2.92 | 2.92 | 2.92 | 2.97 |
| 8' | 7' | 11.89 | 1.44 | 2.55 | 2.52 | 2.52 | 2.52 | 2.52 | 2.52 | 2.52 | 2.49 |
| 3' | 2' | 16.16 | 1.37 | 2.42 | 2.49 | 2.50 | 2.50 | 2.49 | 2.47 | 2.47 | 2.48 |
| 3 | 2a | 18.01 | 1.34 | 2.38 | 2.33 | 2.33 | 2.33 | 2.33 | 2.33 | 2.33 | 2.30 |
| 3 | 6b | 3.59 | 1.76 | 3.11 | 3.32 | 3.38 | 3.15 | 3.09 | 3.14 | 3.14 | 3.39 |
| 3 | 4 | 12.9 | 1.42 | 2.51 | 2.49 | 2.48 | 2.51 | 2.52 | 2.51 | 2.51 | 2.46 |
| 3 | 5b | 14.5 | 1.39 | 2.46 | 2.44 | 2.43 | 2.45 | 2.45 | 2.45 | 2.45 | 2.39 |
| 8 | 3' | 8.2 | 1.53 | 2.71 | 2.75 | 2.72 | 3.83 | 2.33 | 2.19 | 2.19 | 2.45 |
| 8 | 5' | 18.39 | 1.34 | 2.37 | 2.41 | 2.22 | 2.44 | 3.98 | 3.70 | 3.70 | 3.30 |
| 8 | 9 | 11.33 | 1.45 | 2.57 | 3.04 | 3.04 | 2.56 | 3.05 | 2.53 | 2.53 | 2.24 |
| 2a | 2b | 129.71 | 0.97 | 1.71 | 1.77 | 1.77 | 1.77 | 1.77 | 1.77 | 1.77 | 1.75 |
| 2a | 3 | 21.61 | 1.30 | 2.31 | 2.33 | 2.33 | 2.33 | 2.33 | 2.33 | 2.33 | 2.30 |
| 8 | 7a | 6.65 | 1.59 | 2.81 | 3.01 | 3.02 | 2.97 | 2.95 | 2.99 | 2.99 | 3.02 |
| 8 | 7b | 24.92 | 1.27 | 2.25 | 2.31 | 2.31 | 2.29 | 2.30 | 2.31 | 2.31 | 2.29 |
| 6b | 6a | 101.88 | 1.01 | 1.78 | 1.76 | 1.76 | 1.77 | 1.76 | 1.77 | 1.77 | 1.75 |
| 2b | 2a | 142.21 | 0.95 | 1.68 | 1.76 | 1.76 | 1.76 | 1.74 | 1.76 | 1.76 | 1.73 |
| 2b | 3 | 4.46 | 1.69 | 3.00 | 2.93 | 2.92 | 2.97 | 2.98 | 2.97 | 2.97 | 2.88 |
| 6b | 5a | 4.01 | 1.72 | 3.05 | 2.92 | 2.89 | 2.95 | 2.97 | 2.95 | 2.95 | 2.86 |
| 6b | 5b | 19.58 | 1.32 | 2.34 | 2.36 | 2.37 | 2.36 | 2.36 | 2.36 | 2.36 | 2.35 |

## NOE/ROE

**Table A4-35. $\tau_c$ determination. PANIC NOE and ROE intensities for quinine ($\tau_m$ = 300 ms). Corr. factor given by equation 3.14 in the thesis. Corr. rel is the correction by including $\tau_c$ compared to the normal PANIC approach.**

| H1 | H2 | PANIC | | Corr. factor | ROE | NOE/ROE | $\tau_c$ [ps] | Corr Rel. | Rel. NOE | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | NOE | T-ROE | | | | | | NOE | NOE/ROE |
| 2' | 3' | 14.4 | 28.3 | 0.96 | 40.4 | 0.36 | 187.8 | 0.99 | 1.39 | 1.37 |
| 9 | 3' | 5.9 | 10.7 | 1.00 | 15.5 | 0.38 | 180.5 | 0.99 | 1.62 | 1.60 |
| 9 | 5' | 39.3 | 62.2 | 1.00 | 85.2 | 0.46 | 157.4 | 1.01 | 1.18 | 1.19 |
| 9 | 6a | 13.0 | 19.2 | 1.01 | 25.7 | 0.51 | 145.3 | 1.01 | 1.42 | 1.43 |
| 9 | 8 | 11.4 | 19.8 | 1.01 | 28.6 | 0.40 | 175.6 | 1.00 | 1.45 | 1.44 |
| 9 | 7a | 7.4 | 9.5 | 1.04 | 12.2 | 0.60 | 121.5 | 1.02 | 1.56 | 1.58 |
| 6a | 9 | 15.9 | 23.2 | 1.01 | 30.8 | 0.52 | 143.2 | 1.01 | 1.37 | 1.39 |
| 6a | 6b | 105.5 | 180.2 | 0.98 | 250.7 | 0.42 | 168.6 | 1.00 | 1.00 | 1.00 |
| 6a | 5a | 23.5 | 34.8 | 0.99 | 45.5 | 0.52 | 143.2 | 1.01 | 1.28 | 1.30 |
| 8' | 7' | 11.9 | 21.8 | 0.97 | 30.7 | 0.39 | 178.3 | 0.99 | 1.44 | 1.43 |
| 3' | 2' | 16.2 | 31.3 | 0.96 | 44.7 | 0.36 | 186.4 | 0.99 | 1.37 | 1.35 |
| 3 | 2a | 18.0 | 30.5 | 0.98 | 41.9 | 0.43 | 166.0 | 1.00 | 1.34 | 1.34 |
| 3 | 6b | 3.6 | 6.0 | 0.97 | 8.1 | 0.44 | 162.4 | 1.00 | 1.76 | 1.76 |
| 3 | 4 | 12.9 | 20.1 | 0.97 | 26.3 | 0.49 | 149.5 | 1.01 | 1.42 | 1.43 |
| 3 | 5b | 14.5 | 23.5 | 0.96 | 31.2 | 0.46 | 156.4 | 1.01 | 1.39 | 1.40 |
| 8 | 3' | 8.2 | 11.2 | 1.06 | 15.1 | 0.54 | 136.4 | 1.01 | 1.53 | 1.55 |
| 8 | 5' | 18.4 | 31.7 | 1.05 | 47.3 | 0.39 | 177.9 | 0.99 | 1.34 | 1.33 |
| 8 | 9 | 12.3 | 22.1 | 1.01 | 32.4 | 0.38 | 180.3 | 0.99 | 1.43 | 1.42 |
| 2a | 2b | 129.7 | 220.7 | 0.98 | 305.1 | 0.43 | 167.4 | 1.00 | 0.97 | 0.97 |
| 2a | 3 | 21.6 | 35.9 | 0.98 | 49.0 | 0.44 | 162.9 | 1.00 | 1.30 | 1.31 |
| 8 | 7b | 24.9 | 36.1 | 0.98 | 46.3 | 0.54 | 137.3 | 1.01 | 1.27 | 1.29 |
| 6b | 6a | 101.9 | 176.3 | 0.98 | 246.5 | 0.41 | 170.8 | 1.00 | 1.01 | 1.00 |
| 2b | 2a | 142.2 | 252.2 | 0.98 | 354.3 | 0.40 | 174.2 | 1.00 | 0.95 | 0.95 |
| 2b | 3 | 4.5 | 10.0 | 0.96 | 15.0 | 0.30 | 207.3 | 0.97 | 1.69 | 1.64 |
| 6b | 5b | 19.6 | 30.2 | 0.97 | 39.5 | 0.50 | 148.2 | 1.01 | 1.32 | 1.34 |

## RDC

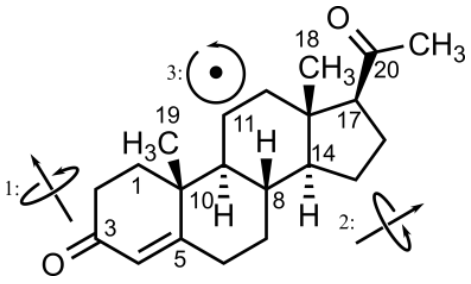
**Table A4-36. RDC extraction of quinine (Polymer 8.6 in the thesis). $^2$H splitting = 49.5 Hz ($CDCl_3$)**

| Quinine | | | | |
|---|---|---|---|---|
| C | H | *J* [Hz] | *T* [Hz] | *D* [Hz] |
| 3 | 3 | 138.2 | 150.5 | 12.3 |
| 4 | 4 | 129.8 | 139.3 | 9.6 |
| 6 | 6a | 140.4 | 113.9 | -26.5 |
| 6 | 6b | 137.1 | 139.1 | 2 |
| 8 | 8 | 135.2 | 132.4 | -2.8 |
| 9 | 9 | 142.2 | 152.3 | 10.1 |
| 2' | 2' | 177.7 | 194 | 16.3 |
| 3' | 3' | 162.8 | 178.2 | 15.5 |
| 5' | 5' | 161.5 | 176.1 | 14.6 |
| 7' | 7' | 158 | 171.2 | 13.2 |
| 8' | 8' | 162.6 | 176.1 | 13.5 |

## A4.3.3. Progesterone

**Table A4-37. NMR assignment of progesterone**

| # | $^{1}$H δ [ppm],mult, *J* [Hz] | $^{13}$C δ[ppm] |
|---|---|---|
| 1a | 2.02 (1H,m) | 35.6 |
| 1b | 1.69 (1H,m) | 35.6 |
| 2a | 2.35 (1H,m) | 33.9 |
| 2b | 2.39 (1H,m) | 33.9 |
| 3 | | 208.8 |
| 4 | 5.72 (1H,s) | 123.9 |
| 5 | | 170.7 |
| 6a | 2.24 (1H,ddd,14.5,3.7,2.4) | 32.7 |
| 6b | 2.38 (1H,m) | 32.7 |
| 7a | 1.06 (1H,m) | 31.9 |
| 7b | 1.95 (1H,m) | 31.9 |
| 8 | 1.55 (1H,m) | 35.5 |
| 9 | 0.98 (1H,m) | 53.7 |
| 10 | | 43.9 |
| 11a | 1.62 (1H,m) | 21 |
| 11b | 1.44 (1H,m) | 21 |
| 12a | 1.43 (1H,m) | 38.6 |
| 12b | 2.03 (1H,m) | 38.6 |
| 13 | | 38.5 |
| 14 | 1.15 (1H,m) | 56 |
| 15a | 1.78 (1H,m) | 24.5 |
| 15b | 1.17 (1H,m) | 24.5 |
| 16a | 1.66 (1H,m) | 22.7 |
| 16b | 2.15 (1H,dt,16.7,3.4) | 22.7 |
| 17 | 2.54 (1H,t,9.1) | 63.4 |
| 18Me | 0.56 (3H,s) | 13.37 |
| 19Me | 1.13 (3H,s) | 17.3 |
| 20 | | 199.3 |
| 21Me | 2.06 (3H,s) | 31.4 |



## NOE/ROE

### $CDCl_3$

**Table A4-38. $\tau_c$ determination. PANIC NOE and ROE intensities for progesterone in $CDCl_3$ ($\tau_m$ = 300 ms). Corr. factor given by equation 3.14 in the thesis. Corr. rel is the correction by including $\tau_c$ compared to the normal PANIC approach.**

| H1 | H2 | PANIC | | Corr. factor | ROE | NOE/ROE | $\tau_c$ [ps] | Corr. rel | Rel. NOE | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | NOE | T-ROE | | | | | | NOE | NOE/ROE |
| 4 | 6a | 16.2 | 16.3 | 1.05 | 17.1 | 0.95 | 36.0 | 1.03 | 1.22 | 1.26 |
| 17 | 16a | 10.3 | 10.4 | 1.00 | 10.5 | 0.98 | 21.4 | 0.95 | 1.32 | 1.26 |
| 17 | 12a | 12.7 | 13.1 | 1.00 | 13.6 | 0.93 | 42.4 | 1.06 | 1.27 | 1.35 |
| 17 | 14 | 15.6 | 16.1 | 1.00 | 16.6 | 0.94 | 39.7 | 1.05 | 1.23 | 1.29 |
| 6a | 4 | 22.4 | 22.5 | 1.05 | 23.7 | 0.94 | 38.3 | 1.04 | 1.16 | 1.21 |
| 6a | 7b | 15.9 | 16.9 | 0.99 | 17.8 | 0.89 | 53.7 | 1.09 | 1.23 | 1.34 |
| 6a | 7a | 10.2 | 10.7 | 1.00 | 11.1 | 0.92 | 46.4 | 1.07 | 1.32 | 1.41 |
| 16b | 16a | 55.6 | 60.1 | 0.99 | 64.0 | 0.87 | 59.8 | 1.10 | 0.99 | 1.10 |
| 16b | 15b | 14.5 | 17.7 | 0.98 | 20.4 | 0.71 | 96.8 | 1.15 | 1.24 | 1.43 |
| 7b | 6b | 9.9 | 10.4 | 1.00 | 10.8 | 0.92 | 44.9 | 1.06 | 1.33 | 1.41 |
| 7b | 6a | 14.2 | 15.4 | 0.99 | 16.5 | 0.86 | 63.1 | 1.11 | 1.25 | 1.39 |
| 7b | 8 | 6.9 | 7.8 | 0.98 | 8.6 | 0.80 | 75.9 | 1.13 | 1.41 | 1.59 |
| 7b | 15b | 7.1 | 7.7 | 0.99 | 8.3 | 0.85 | 64.5 | 1.11 | 1.40 | 1.56 |
| 7b | 7a | 58.6 | 64.1 | 0.99 | 68.7 | 0.85 | 64.2 | 1.11 | 0.99 | 1.10 |
| 8 | 6b | 9.8 | 10.1 | 1.00 | 10.3 | 0.95 | 37.2 | 1.04 | 1.33 | 1.38 |
| 8 | 7b | 5.7 | 6.6 | 0.98 | 7.4 | 0.76 | 85.8 | 1.14 | 1.46 | 1.66 |
| 15b | 16b | 15.7 | 17.7 | 0.99 | 19.5 | 0.80 | 76.2 | 1.13 | 1.23 | 1.39 |
| 15b | 8 | 6.9 | 7.5 | 0.99 | 7.9 | 0.86 | 61.1 | 1.11 | 1.41 | 1.56 |
| 7a | 6a | 7.8 | 7.9 | 1.00 | 8.0 | 0.97 | 27.3 | 0.99 | 1.38 | 1.37 |
| 7a | 7b | 54.0 | 54.9 | 1.00 | 55.9 | 0.97 | 29.2 | 1.00 | 1.00 | 1.00 |
| 7a | 15a | 4.8 | 5.7 | 0.98 | 6.5 | 0.73 | 91.8 | 1.15 | 1.50 | 1.72 |
| 9 | 1b | 19.5 | 20.3 | 0.99 | 21.0 | 0.93 | 41.9 | 1.05 | 1.18 | 1.25 |

### DMSO-$d_6$

**Table A4-39. $\tau_c$ determination. PANIC NOE and ROE intensities for progesterone in DMSO-$d_6$ ($\tau_m$ = 300 ms). Corr. factor given by equation 3.14 in the thesis. Corr. rel is the correction by including $\tau_c$ compared to the normal PANIC approach.**

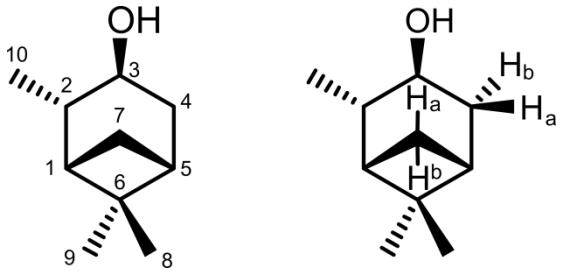
| H1 | H2 | PANIC | | Corr. factor | ROE | NOE/ROE | $\tau_c$ [ps] | Corr. rel | Rel. NOE | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | NOE | T-ROE | | | | | | NOE | NOE/ROE |
| 4 | 6a | 27.7 | 31.2 | 1.04 | 36.2 | 0.77 | 84.8 | 0.99 | 1.27 | 1.26 |
| 17 | 12b | 6.7 | 7.7 | 0.99 | 8.6 | 0.77 | 82.5 | 0.99 | 1.61 | 1.59 |
| 17 | 16a | 20.0 | 24.0 | 0.99 | 27.7 | 0.72 | 94.6 | 1.00 | 1.34 | 1.34 |
| 17 | 12a | 21.3 | 26.7 | 0.99 | 31.6 | 0.68 | 105.2 | 1.00 | 1.33 | 1.33 |
| 17 | 14 | 25.1 | 36.5 | 0.98 | 46.7 | 0.54 | 137.6 | 1.00 | 1.29 | 1.30 |
| 6b | 6a | 117.8 | 134.6 | 0.99 | 150.1 | 0.78 | 80.2 | 0.99 | 1.00 | 0.99 |
| 2b | 2a | 123.6 | 162.5 | 0.98 | 197.9 | 0.62 | 116.7 | 1.01 | 0.99 | 1.00 |
| 2b | 1a | 14.6 | 20.9 | 0.98 | 26.6 | 0.55 | 135.2 | 1.00 | 1.42 | 1.42 |
| 6b | 7b | 18.8 | 22.7 | 0.99 | 26.3 | 0.72 | 95.7 | 1.00 | 1.36 | 1.35 |
| 6b | 8 | 15.3 | 20.7 | 0.98 | 25.6 | 0.60 | 123.2 | 1.01 | 1.40 | 1.41 |
| 6b | 7a | 4.2 | 6.0 | 0.97 | 7.6 | 0.55 | 134.9 | 1.00 | 1.74 | 1.75 |
| 6a | 7b | 17.3 | 25.8 | 0.97 | 33.4 | 0.52 | 142.8 | 1.00 | 1.38 | 1.38 |
| 6a | 7a | 19.7 | 24.3 | 0.98 | 28.4 | 0.69 | 101.6 | 1.00 | 1.35 | 1.35 |
| 2a | 1a | 11.3 | 18.4 | 0.97 | 24.7 | 0.46 | 158.1 | 1.00 | 1.48 | 1.47 |
| 2a | 1b | 17.5 | 23.9 | 0.98 | 29.7 | 0.59 | 125.5 | 1.01 | 1.37 | 1.38 |
| 7b | 6b | 20.3 | 24.6 | 0.99 | 28.5 | 0.71 | 96.7 | 1.00 | 1.34 | 1.34 |
| 7b | 6a | 17.3 | 24.7 | 0.98 | 31.3 | 0.55 | 133.8 | 1.00 | 1.38 | 1.38 |
| 7b | 15b | 13.3 | 21.1 | 0.97 | 27.8 | 0.48 | 152.7 | 1.00 | 1.44 | 1.43 |
| 7b | 8 | 13.5 | 21.6 | 0.96 | 28.6 | 0.47 | 154.3 | 1.00 | 1.43 | 1.43 |
| 7b | 14 | 13.3 | 18.9 | 0.97 | 23.6 | 0.56 | 131.5 | 1.00 | 1.44 | 1.44 |
| 7b | 7a | 117.1 | 143.4 | 0.98 | 165.8 | 0.71 | 98.2 | 1.00 | 1.00 | 1.00 |

## A4.4. RDC compounds

### A4.4.1. IPC

**Table A4-40. NMR assignment of IPC**

| # | $\delta_H$ [ppm] (int,mult,*J* [Hz]) | $\delta_C$ [ppm] |
|---|---|---|
| 1 | 1.78 (1H, td, 5.9,1.4) | 47.72 |
| 2 | 1.94 (1H, m) | 47.56 |
| 3 | 4.06 (1H, dt, 9.5,4.8) | 69.8 |
| 4a | 2.50 (1H, dddd,13.4,10.0,3.1,2.7) | 38.9 |
| 4b | 1.69 (1H, ddd, 13.9,4.5,2.6) | |
| 5 | 1.93 (1H, m) | 41.8 |
| 6 | | 39.2 |
| 7a | 1.03 (1H, d, 10.0) | 34.1 |
| 7b | 2.36 (1H, dtd, 9.4,6.4,2.7) | |
| 8 | 1.22 (3H, s) | 28.0 |
| 9 | 0.94 (3H, s) | 24.0 |
| 10 | 1.11 (3H, d, 7.2) | 21.1 |
| OH | 4.63 | |




### RDC

**Table A4-41. RDC extraction of IPC (Polymer 6.1 in the thesis). $^2$H splitting = 44.2 Hz ($CDCl_3$)**

| **IPC** | | | | |
|---|---|---|---|---|
| Nuc 1 | Nuc 2 | *J* [Hz] | *T* [Hz] | *D* [Hz] |
| C1 | H1 | 140.4 | 156.7 | 16.3 |
| C2 | H2 | 126.6 | 114.9 | -11.6 |
| C3 | H3 | 142.3 | 156.9 | 14.6 |
| C4 | H4a | 126.9 | 122.2 | -4.8 |
| C4 | H4b | 126.6 | 142.1 | 15.5 |
| C5 | H5 | 141.4 | 134 | -7.4 |
| C7 | H7a | 147.2 | 125.5 | -21.7 |
| C7 | H7b | 135.2 | 140.2 | 4.9 |
| H1 | H5 | 2.5 | 3.5 | 1 |
| H1 | H7a | 1.8 | 2.3 | 0.5 |
| H2 | H3 | 5.3 | 9.0 | 3.7 |
| H2 | H4b | -0.7 | 0.8 | 1.5 |
| H3 | H4a | 4.7 | 6.3 | 1.6 |
| H3 | H4b | 9.2 | 7.1 | -2.1 |
| H3 | H5 | -0.2 | -0.5 | -0.3 |
| H4a | H5 | 2.8 | -4.2 | -7 |

**Table A4-42. RDCs in Hz back-calculated from different assignments of IPC using SVD (Polymer 6.1 in the thesis).**

| Nuc 1 | Nuc 2 | Exp | Right | Dia-2 | Dia-3 | Dia-2+3 | Dia-2+4+7 | Dia-2+3+4+8 |
|---|---|---|---|---|---|---|---|---|
| C1 | H1 | 16.3 | 16.7 | 20.7 | 16.2 | 17.7 | 20.5 | 16.4 |
| C2 | H2 | -11.6 | -9.6 | -7.8 | -4.0 | -14.3 | -6.4 | -16.3 |
| C3 | H3 | 14.6 | 14.2 | -0.7 | 10.7 | -1.9 | 9.6 | 7.9 |
| C4 | H4a | -4.8 | -4.8 | -11.2 | -10.5 | -9.4 | -8.2 | -0.4 |
| C4 | H4b | 15.5 | 14.2 | 7.3 | 18.2 | 8.9 | 15.9 | 14.7 |
| C5 | H6 | -7.4 | -7.8 | -7.9 | -10.1 | -8.6 | -5.1 | -5.4 |
| C7 | H7a | -21.7 | -23.2 | -6.1 | -18.2 | -5.5 | -18.3 | -19.7 |
| C7 | H7b | 4.9 | 5.2 | 6.4 | 5.8 | 5.0 | 9.0 | 9.4 |
| | | *Q*-factor | 0.076 | 0.661 | 0.305 | 0.655 | 0.284 | 0.283 |

**Table A4-43. RDCs in Hz back-calculated from different assignments of IPC using SVD (Polymer 6.1 in the thesis).**

| Nuc 1 | Nuc 2 | Exp | Right | Dia-2 | Dia-3 | Dia-2+3 | Dia-2+4+7 | Dia-2+3+4+8 |
|---|---|---|---|---|---|---|---|---|
| C1 | H1 | 16.3 | 18.2 | 14.0 | 13.6 | 22.5 | 16.5 | 13.0 |
| C2 | H2 | -11.6 | -8.8 | -6.0 | 2.6 | -7.7 | -4.3 | -4.1 |
| C3 | H3 | 14.6 | 12.5 | -11.0 | 11.5 | -8.2 | 12.4 | 11.2 |
| C4 | H4a | -4.8 | -6.1 | -6.0 | -6.8 | -6.5 | -8.6 | -2.9 |
| C4 | H4b | 15.5 | 16.1 | 8.1 | 15.6 | 6.3 | 12.7 | 9.9 |
| C5 | H6 | -7.4 | -5.8 | -0.7 | -5.3 | -3.5 | -1.8 | -0.5 |
| C7 | H7a | -21.7 | -23.8 | -11.2 | -19.0 | -11.0 | -16.9 | -18.9 |
| C7 | H7b | 4.9 | 2.7 | -1.7 | 0.9 | -0.7 | 10.8 | 11.2 |
| H1 | H6 | 1.0 | 0.9 | 1.0 | 0.7 | 1.2 | 1.4 | 1.3 |
| H1 | H7a | 0.5 | 0.0 | 0.0 | -0.7 | 0.8 | -0.3 | -1.0 |
| H2 | H3 | 3.7 | 4.0 | 6.3 | 7.5 | 2.7 | 3.5 | 5.0 |
| H2 | H4b | 1.5 | 0.8 | 0.7 | 0.7 | 1.1 | 1.9 | 1.0 |
| H3 | H4a | 1.6 | 1.1 | -0.9 | 0.7 | -2.4 | 0.2 | 1.4 |
| H3 | H4b | -2.1 | -1.0 | -0.8 | -0.5 | -1.4 | -2.6 | -2.3 |
| H3 | H6 | -0.3 | -0.5 | -0.6 | -0.3 | -0.6 | -0.5 | -0.5 |
| H4a | H6 | -7.0 | -7.9 | -2.8 | -6.3 | -1.6 | -6.2 | -6.8 |
| | | *Q*-factor | 0.148 | 0.809 | 0.426 | 0.762 | 0.340 | 0.377 |

## Chiral polymers

**Table A4-44. Chirality of IPC and polymers in the following tables**

| | L-TEGPhe/DMAA | D-TEGPhe/DMAA |
|---|---|---|
| **(+)-IPC** | 1,2,3 | 9,10 |
| **(-)-IPC** | 4,5,6 | 7,8 |

**Table A4-45. Output from MSPIN upon entering the extracted RDCs of IPC aligned in the L/D-TEG-Phe/DMAA**

| **1** | **Exp Hz** | **Comp Hz** | **2** | **Exp Hz** | **Comp Hz** | **3** | **Exp Hz** | **Comp Hz** |
|---|---|---|---|---|---|---|---|---|
| C3,H3 | -12 | -11.1 | C3,H3 | -13 | -12.28 | C3,H3 | -14 | -12.66 |
| C1,H1 | -9.9 | -9.49 | C1,H1 | -9 | -7.85 | C1,H1 | -8.7 | -7.42 |
| C2,H2 | 0.6 | -0.86 | C2,H2 | -2.1 | -2.12 | C2,H2 | -1.4 | -1.71 |
| C5,H6 | -0.3 | 0.73 | C5,H6 | -1.6 | 0.03 | C5,H6 | -1.5 | -0.22 |
| C4,H4 | -17.2 | -16.1 | C4,H4 | -16.8 | -16.07 | C4,H4 | -15.7 | -15.62 |
| C4,H5 | 3.1 | 3.22 | C4,H5 | 1.3 | 2.08 | C4,H5 | 2.3 | 2.07 |
| C7,H8 | 5.5 | 5.74 | C7,H8 | 5.8 | 7.07 | C7,H8 | 6.1 | 6.58 |
| C7,H7 | 13.5 | 15.32 | C7,H7 | 13.5 | 14.8 | C7,H7 | 13.5 | 15.16 |
| C9,H9 | -4.64 | -5.67 | C9,H9 | -6.43 | -6.14 | C9,H9 | -5.44 | -6.11 |
| C8,H8 | 3.43 | 3.55 | C8,H8 | 4.31 | 3.91 | C8,H8 | 4.64 | 3.99 |
| C10,H10 | 1.25 | 1.49 | C10,H10 | 2.18 | 1.92 | C10,H10 | 2.57 | 2.11 |
| *Q*-factor | | 0.112 | *Q*-factor | | 0.106 | *Q*-factor | | 0.107 |
| Conformer 1 | | | Conformer 1 | | | Conformer 1 | | |
| Alignment tensor | | | Alignment tensor | | | Alignment tensor | | |
| A'x= 7.648e-05 | | | A'x= 9.749e-05 | | | A'x= 1.049e-04 | | |
| A'y= 2.528e-04 | | | A'y= 2.439e-04 | | | A'y= 2.381e-04 | | |
| A'z=-3.292e-04 | | | A'z=-3.414e-04 | | | A'z=-3.431e-04 | | |
| Saupe tensor | | | Saupe tensor | | | Saupe tensor | | |
| S'x= 1.147e-04 | | | S'x= 1.462e-04 | | | S'x= 1.574e-04 | | |
| S'y= 3.791e-04 | | | S'y= 3.659e-04 | | | S'y= 3.572e-04 | | |
| S'z=-4.939e-04 | | | S'z=-5.122e-04 | | | S'z=-5.146e-04 | | |
| Alignment tensor eigenvectors | | | Alignment tensor eigenvectors | | | Alignment tensor eigenvectors | | |
| e[x]=(-0.000, 0.040,-0.999) | | | e[x]=(-0.035, 0.103,-0.994) | | | e[x]=(-0.043, 0.146,-0.988) | | |
| e[y]=(-0.503, 0.863, 0.035) | | | e[y]=(-0.468, 0.877, 0.107) | | | e[y]=(-0.472, 0.869, 0.149) | | |
| e[z]=( 0.864, 0.503, 0.020) | | | e[z]=( 0.883, 0.469, 0.018) | | | e[z]=( 0.881, 0.472, 0.031) | | |
| Alignment tensor in laboratory coordinates: | | | Alignment tensor in laboratory coordinates: | | | Alignment tensor in laboratory coordinates: | | |
| [-1.818e-04,-2.529e-04,-1.005e-05] | | | [-2.129e-04,-2.417e-04,-1.429e-05] | | | [-2.130e-04,-2.410e-04,-2.175e-05] | | |
| [-2.529e-04,1.053e-04,1.248e-06] | | | [-2.417e-04,1.139e-04,1.014e-05] | | | [-2.410e-04,1.055e-04,1.065e-05] | | |
| [-1.005e-05,1.248e-06,7.653e-05] | | | [-1.429e-05,1.014e-05,9.904e-05] | | | [-2.175e-05,1.065e-05,1.075e-04] | | |
| SVD condition number is 2.411e+00 | | | SVD condition number is 2.411e+00 | | | SVD condition number is 2.411e+00 | | |
| Axial component Aa = -4.939e-04 | | | Axial component Aa = -5.122e-04 | | | Axial component Aa = -5.146e-04 | | |
| Rhombic component Ar = -1.763e-04 | | | Rhombic component Ar = -1.465e-04 | | | Rhombic component Ar = -1.332e-04 | | |
| rhombicity R = 3.570e-01 | | | rhombicity R = 2.860e-01 | | | rhombicity R = 2.588e-01 | | |
| Asimmetry parameter etha =5.354e-01 | | | Asimmetry parameter etha =4.290e-01 | | | Asimmetry parameter etha =3.882e-01 | | |
| GDO = 6.098e-04 | | | GDO = 6.180e-04 | | | GDO = 6.162e-04 | | |
| Euler Angles (degrees) | | | Euler Angles (degrees) | | | Euler Angles (degrees) | | |
| Set 1 | | | Set 1 | | | Set 1 | | |
| (87.7,-59.8,-90.0) | | | (87.8,-62.0,-94.2) | | | (86.2,-61.7,-95.2) | | |
| Set 2 | | | Set 2 | | | Set 2 | | |
| (-92.3,239.8,90.0) | | | (-92.2,242.0,85.8) | | | (-93.8,241.7,84.8) | | |

| **4** | **Exp Hz** | **Comp Hz** | **5** | **Exp Hz** | **Comp Hz** | **6** | **Exp Hz** | **Comp Hz** |
|---|---|---|---|---|---|---|---|---|
| C3,H3 | -12.3 | -10.41 | C3,H3 | -12.6 | -10.65 | C3,H3 | -12.5 | -10.84 |
| C1,H1 | -10.1 | -8.44 | C1,H1 | -11.5 | -10.51 | C1,H1 | -10.4 | -9.84 |
| C2,H2 | -0.7 | -0.08 | C2,H2 | 2.8 | 1.3 | C2,H2 | 2.9 | 1.01 |
| C5,H6 | -1.3 | -0.83 | C5,H6 | -1.2 | -0.33 | C5,H6 | -1.2 | -0.52 |
| C4,H4 | -11.7 | -13.44 | C4,H4 | -13.3 | -13.24 | C4,H4 | -13.8 | -13.36 |
| C4,H5 | 3.1 | 1.41 | C4,H5 | 1.4 | 0.2 | C4,H5 | 1.7 | 0.67 |
| C7,H8 | 7.2 | 6.72 | C7,H8 | 8 | 7.45 | C7,H8 | 7.7 | 7.05 |
| C7,H7 | 13.5 | 14.64 | C7,H7 | 13.5 | 15.93 | C7,H7 | 13.5 | 15.84 |
| C9,H9 | -5.61 | -5.73 | C9,H9 | -5.56 | -6.15 | C9,H9 | -5.02 | -6.06 |
| C8,H8 | 3.3 | 3.28 | C8,H8 | 4.44 | 3.33 | C8,H8 | 3.37 | 3.38 |
| C10,H10 | 1.54 | 2.23 | C10,H10 | 2.66 | 2.68 | C10,H10 | 2.33 | 2.6 |
| *Q*-factor | | 0.148 | *Q*-factor | | 0.147 | *Q*-factor | | 0.143 |
| Conformer 1 | | | Conformer 1 | | | Conformer 1 | | |
| Alignment tensor | | | Alignment tensor | | | Alignment tensor | | |
| A'x= 1.119e-04 | | | A'x= 1.328e-04 | | | A'x= 1.299e-04 | | |
| A'y= 2.115e-04 | | | A'y= 2.214e-04 | | | A'y= 2.184e-04 | | |
| A'z=-3.234e-04 | | | A'z=-3.542e-04 | | | A'z=-3.483e-04 | | |
| Saupe tensor | | | Saupe tensor | | | Saupe tensor | | |
| S'x= 1.678e-04 | | | S'x= 1.991e-04 | | | S'x= 1.949e-04 | | |
| S'y= 3.173e-04 | | | S'y= 3.321e-04 | | | S'y= 3.275e-04 | | |
| S'z=-4.851e-04 | | | S'z=-5.313e-04 | | | S'z=-5.224e-04 | | |
| Alignment tensor eigenvectors | | | Alignment tensor eigenvectors | | | Alignment tensor eigenvectors | | |
| e[x]=(-0.038, 0.034, 0.999) | | | e[x]=(-0.092, 0.178, 0.980) | | | e[x]=(-0.066, 0.111, 0.992) | | |
| e[y]=( 0.489,-0.871, 0.048) | | | e[y]=( 0.510,-0.837, 0.200) | | | e[y]=( 0.508,-0.851, 0.129) | | |
| e[z]=( 0.872, 0.490, 0.017) | | | e[z]=( 0.855, 0.518,-0.014) | | | e[z]=( 0.859, 0.512, 0.000) | | |
| Alignment tensor in laboratory coordinates: | | | Alignment tensor in laboratory coordinates: | | | Alignment tensor in laboratory coordinates: | | |
| [-1.950e-04,-2.283e-04,-3.949e-06] | | | [-2.005e-04,-2.535e-04,1.473e-05] | | | [-1.998e-04,-2.487e-04,5.763e-06] | | |
| [-2.283e-04,8.300e-05,-7.673e-06] | | | [-2.535e-04,6.426e-05,-1.136e-05] | | | [-2.487e-04,6.845e-05,-9.774e-06] | | |
| [-3.949e-06,-7.673e-06,1.120e-04] | | | [ 1.473e-05,-1.136e-05,1.362e-04] | | | [ 5.763e-06,-9.774e-06,1.314e-04] | | |
| SVD condition number is 2.411e+00 | | | SVD condition number is 2.411e+00 | | | SVD condition number is 2.411e+00 | | |
| Axial component Aa = -4.851e-04 | | | Axial component Aa = -5.313e-04 | | | Axial component Aa = -5.224e-04 | | |
| Rhombic component Ar = -9.966e-05 | | | Rhombic component Ar = -8.866e-05 | | | Rhombic component Ar = -8.845e-05 | | |
| rhombicity R = 2.054e-01 | | | rhombicity R = 1.669e-01 | | | rhombicity R = 1.693e-01 | | |
| Asimmetry parameter etha =3.081e-01 | | | Asimmetry parameter etha =2.503e-01 | | | Asimmetry parameter etha =2.540e-01 | | |
| GDO = 5.733e-04 | | | GDO = 6.230e-04 | | | GDO = 6.129e-04 | | |
| Euler Angles (degrees) | | | Euler Angles (degrees) | | | Euler Angles (degrees) | | |
| Set 1 | | | Set 1 | | | Set 1 | | |
| (88.1,-60.6,94.4) | | | (91.5,-58.8,100.2) | | | (90.0,-59.2,97.4) | | |
| Set 2 | | | Set 2 | | | Set 2 | | |
| (-91.9,240.6,-85.6) | | | (-88.5,238.8,-79.8) | | | (-90.0,239.2,-82.6) | | |

| **7** | **Exp Hz** | **Comp Hz** | **8** | **Exp Hz** | **Comp Hz** |
|---|---|---|---|---|---|
| C3,H3 | -6.5 | -5.74 | C3,H3 | -6 | -5.68 |
| C1,H1 | -5.4 | -4.42 | C1,H1 | -4.5 | -3.99 |
| C2,H2 | -0.9 | -0.18 | C2,H2 | -1.4 | -1.05 |
| C5,H6 | -1.7 | -1.24 | C5,H6 | -2 | -1.62 |
| C4,H4 | -6.9 | -7.71 | C4,H4 | -8.4 | -8.62 |
| C4,H5 | 1.7 | 1.24 | C4,H5 | 2.1 | 2.09 |
| C7,H8 | 3.8 | 3.98 | C7,H8 | 3.9 | 4.19 |
| C7,H7 | 8.1 | 8.47 | C7,H7 | 8.1 | 8.32 |
| C9,H9 | -2.7 | -3.3 | C9,H9 | -2.6 | -3.34 |
| C8,H8 | 2.1 | 1.81 | C8,H8 | 2.4 | 1.84 |
| C10,H10 | 1.2 | 1.3 | C10,H10 | 1.2 | 1.07 |
| *Q*-factor | | 0.131 | *Q*-factor | | 0.085 |
| Conformer 1 | | | Conformer 1 | | |
| Alignment tensor | | | Alignment tensor | | |
| A'x= 6.359e-05 | | | A'x= 5.180e-05 | | |
| A'y= 1.205e-04 | | | A'y= 1.314e-04 | | |
| A'z=-1.841e-04 | | | A'z=-1.832e-04 | | |
| Saupe tensor | | | Saupe tensor | | |
| S'x= 9.538e-05 | | | S'x= 7.771e-05 | | |
| S'y= 1.808e-04 | | | S'y= 1.970e-04 | | |
| S'z=-2.761e-04 | | | S'z=-2.747e-04 | | |
| Alignment tensor eigenvectors | | | Alignment tensor eigenvectors | | |
| e[x]=(-0.086, 0.063, 0.994) | | | e[x]=(-0.094, 0.020, 0.995) | | |
| e[y]=( 0.461,-0.882, 0.096) | | | e[y]=( 0.428,-0.902, 0.059) | | |
| e[z]=( 0.883, 0.467, 0.047) | | | e[z]=( 0.899, 0.432, 0.076) | | |
| Alignment tensor in laboratory coordinates: | | | Alignment tensor in laboratory coordinates: | | |
| [-1.174e-04,-1.253e-04,-7.703e-06] | | | [-1.234e-04,-1.219e-04,-1.399e-05] | | |
| [-1.253e-04,5.387e-05,-1.024e-05] | | | [-1.219e-04,7.270e-05,-1.189e-05] | | |
| [-7.703e-06,-1.024e-05,6.357e-05] | | | [-1.399e-05,-1.189e-05,5.073e-05] | | |
| SVD condition number is 2.411e+00 | | | SVD condition number is 2.411e+00 | | |
| Axial component Aa = -2.761e-04 | | | Axial component Aa = -2.747e-04 | | |
| Rhombic component Ar = -5.691e-05 | | | Rhombic component Ar = -7.956e-05 | | |
| rhombicity R = 2.061e-01 | | | rhombicity R = 2.896e-01 | | |
| Asimmetry parameter etha =3.092e-01 | | | Asimmetry parameter etha =4.343e-01 | | |
| GDO = 3.264e-04 | | | GDO = 3.319e-04 | | |
| Euler Angles (degrees) | | | Euler Angles (degrees) | | |
| Set 1 | | | Set 1 | | |
| (84.3,-62.0,100.6) | | | (80.1,-64.0,102.3) | | |
| Set 2 | | | Set 2 | | |
| (-95.7,242.0,-79.4) | | | (-99.9,244.0,-77.7) | | |

| **9** | **Exp Hz** | **Comp Hz** | **10** | **Exp Hz** | **Comp Hz** |
|---|---|---|---|---|---|
| C3,H3 | -5.8 | -5.01 | C3,H3 | -6.3 | -5.42 |
| C1,H1 | -6.2 | -5.36 | C1,H1 | -6.1 | -5.45 |
| C2,H2 | -0.5 | 0.28 | C2,H2 | -0.4 | 0.35 |
| C5,H6 | -1 | -0.96 | C5,H6 | 0.5 | -0.02 |
| C4,H4 | -6 | -7.23 | C4,H4 | -5.4 | -7.15 |
| C4,H5 | 1.5 | 0.6 | C4,H5 | 1.8 | 0.27 |
| C7,H8 | 4.7 | 4.44 | C7,H8 | 4.8 | 3.91 |
| C7,H7 | 8.1 | 8.23 | C7,H7 | 8.1 | 8.02 |
| C9,H9 | -2.4 | -3.28 | C9,H9 | -2.3 | -3.15 |
| C8,H8 | 2 | 1.6 | C8,H8 | 1.6 | 1.71 |
| C10,H10 | 0.9 | 1.32 | C10,H10 | 0.8 | 1.25 |
| *Q*-factor | | 0.16 | *Q*-factor | | 0.21 |
| Conformer 1 | | | Conformer 1 | | |
| Alignment tensor | | | Alignment tensor | | |
| A'x= 6.133e-05 | | | A'x= 6.263e-05 | | |
| A'y= 1.212e-04 | | | A'y= 1.177e-04 | | |
| A'z=-1.826e-04 | | | A'z=-1.803e-04 | | |
| Saupe tensor | | | Saupe tensor | | |
| S'x= 9.200e-05 | | | S'x= 9.395e-05 | | |
| S'y= 1.819e-04 | | | S'y= 1.766e-04 | | |
| S'z=-2.739e-04 | | | S'z=-2.705e-04 | | |
| Alignment tensor eigenvectors | | | Alignment tensor eigenvectors | | |
| e[x]=(-0.134, 0.236, 0.963) | | | e[x]=(-0.063, 0.143, 0.988) | | |
| e[y]=( 0.466,-0.842, 0.271) | | | e[y]=( 0.508,-0.847, 0.155) | | |
| e[z]=( 0.875, 0.485, 0.003) | | | e[z]=( 0.859, 0.511,-0.020) | | |
| Alignment tensor in laboratory coordinates: | | | Alignment tensor in laboratory coordinates: | | |
| [-1.122e-04,-1.269e-04,6.895e-06] | | | [-1.025e-04,-1.305e-04,8.406e-06] | | |
| [-1.269e-04,4.652e-05,-1.404e-05] | | | [-1.305e-04,3.863e-05,-4.787e-06] | | |
| [ 6.895e-06,-1.404e-05,6.573e-05] | | | [ 8.406e-06,-4.787e-06,6.386e-05] | | |
| SVD condition number is 2.411e+00 | | | SVD condition number is 2.411e+00 | | |
| Axial component Aa = -2.739e-04 | | | Axial component Aa = -2.705e-04 | | |
| Rhombic component Ar = -5.990e-05 | | | Rhombic component Ar = -5.507e-05 | | |
| rhombicity R = 2.187e-01 | | | rhombicity R = 2.036e-01 | | |
| Asimmetry parameter etha =3.281e-01 | | | Asimmetry parameter etha =3.054e-01 | | |
| GDO = 3.246e-04 | | | GDO = 3.195e-04 | | |
| Euler Angles (degrees) | | | Euler Angles (degrees) | | |
| Set 1 | | | Set 1 | | |
| (89.6,-61.0,106.0) | | | (92.2,-59.2,97.0) | | |
| Set 2 | | | Set 2 | | |
| (-90.4,241.0,-74.0) | | | (-87.8,239.2,-83.0) | | |

**Table A4-46. Comparison of the alignment frames obtained from MSPIN upon entering the extracted RDCs of IPC aligned in the L- or D-TEG-Phe/DMAA**

| GCB | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0.995 | 0.993 | 0.991 | 0.981 | 0.985 | 0.987 | 0.982 | 0.982 | 0.987 |
| 2 | | 1 | 0.999 | 0.995 | 0.982 | 0.987 | 0.993 | 0.99 | 0.985 | 0.986 |
| 3 | | | 1 | 0.995 | 0.983 | 0.988 | 0.994 | 0.989 | 0.984 | 0.986 |
| 4 | | | | 1 | 0.995 | 0.998 | 0.998 | 0.986 | 0.996 | 0.997 |
| 5 | | | | | 1 | 0.999 | 0.988 | 0.967 | 0.996 | 0.999 |
| 6 | | | | | | 1 | 0.992 | 0.973 | 0.996 | 0.999 |
| 7 | | | | | | | 1 | 0.994 | 0.995 | 0.99 |
| 8 | | | | | | | | 1 | 0.981 | 0.971 |
| 9 | | | | | | | | | 1 | 0.997 |
| 10 | | | | | | | | | | 1 |

**Table A4-47. Comparison of the alignment frames obtained from MSPIN upon entering the extracted RDCs of IPC aligned in the L- or D-TEG-Phe/DMAA**

| GCB | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 5.89 | 6.85 | 7.6 | 11.06 | 9.87 | 9.31 | 10.77 | 10.76 | 9.36 |
| 2 | | 0 | 2.12 | 5.91 | 10.91 | 9.39 | 6.58 | 7.94 | 9.88 | 9.56 |
| 3 | | | 0 | 5.64 | 10.7 | 9.01 | 6.21 | 8.34 | 10.13 | 9.71 |
| 4 | | | | 0 | 5.57 | 3.92 | 4.01 | 9.61 | 4.87 | 4.75 |
| 5 | | | | | 0 | 1.87 | 8.83 | 14.82 | 5.29 | 2.17 |
| 6 | | | | | | 0 | 7.29 | 13.34 | 4.99 | 2.46 |
| 7 | | | | | | | 0 | 6.39 | 5.85 | 8.26 |
| 8 | | | | | | | | 0 | 11.06 | 13.75 |
| 9 | | | | | | | | | 0 | 4.75 |
| 10 | | | | | | | | | | 0 |

## A4.4.2. Menthol

**Table A4-48. NMR assignment of menthol in $CDCl_3$**

| # | $\delta_H$ [ppm] (int,mult,*J* [Hz]) | $\delta_C$ [ppm] |
|---|---|---|
| 1 | 3.39 (1H,td,10.5,4.3) | 71.4 |
| 2 | 1.09 (1H,12.4,9.8,2.8) | 50.1 |
| 3a | 0.95 (1H,m) | |
| 3e | 1.59 (1H,dq,12.9.3.1) | 22.9 |
| 4a | 0.83 (1H,m) | |
| 4e | 1.64 (1H,m) | 34.4 |
| 5 | 1.40 (1H,m) | 31.6 |
| 6a | 0.93 (1H,m) | |
| 6e | 1.95 (1H,dtd,12.0,3.8,1.7) | 44.8 |
| 7 | 0.91 (3H,d,7.2) | 22.2 |
| 8 | 2.16 (1H,sept-d,7.0.2.8) | 25.7 |
| 9 | 0.79 (3H,d,7.0) | 15.8 |
| 10 | 0.89 (3H,d,6.4) | 21 |
| OH | 1.6 (1H,br. s) | |

**Table A4-50. NMR assignment of menthol in DMSO-$d_6$**

| # | $\delta_H$ [ppm] (int,mult,*J* [Hz]) | $\delta_C$ [ppm] |
|---|---|---|
| 1 | 3.15 (1H, ddd, 15.2,10.1,5.1) | 69.3 |
| 2 | 0.97 (1H, m) | 49.3 |
| 3a | 0.88 (1H, m) | 22.5 |
| 3e | 1.50 (1H, dq, 12.4,3.0) | 22.5 |
| 4a | 0.75 (1H, m) | 34.1 |
| 4e | 1.59 (1H, dq, 12.3,2.3) | 34.1 |
| 5 | 1.34 (1H,m) | 30.9 |
| 6a | 0.84 (1H, m) | 45 |
| 6e | 1.81 (1H, dq, 12.2, 4.6) | 45 |
| 7 | 0.85 (3H, d, 7.0) | 22 |
| 8 | 2.18 (1H, sept-d, 7.0,2.4) | 24.7 |
| 9 | 0.72 (3H, d, 7.1) | 15.7 |
| 10 | 0.84 (3H, d, 7.0) | 20.7 |
| OH | 4.28 (1H,br. s) | |

### Chiral polymers

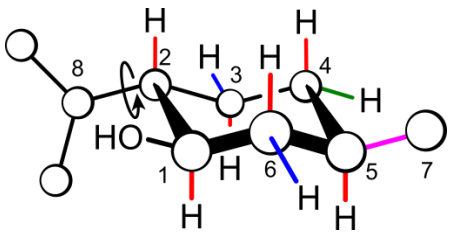
**Table A4-49. Comparison of extracted RDCs ($^1D_{CH}$) in Hz of menthol aligned in the L/D-TEG-Phe/DMAA (1,2 = (+) and 3,4 = (-))**

| H | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 0.4 | 0.4 | 0.5 | 0.4 |
| 2 | 2.6 | 2.6 | 2.6 | 2.6 |
| 6e | 4.7 | 0.5 | 0.4 | 4.5 |
| 6a | -4.3 | -2.5 | -1.8 | -0.9 |
| 4e | 3.1 | 1.8 | 0.2 | 1.8 |
| 4a | -10.4 | -7.4 | -4.4 | -6.9 |
| 5 | 0.3 | 1.9 | 0.8 | 1.0 |
| 7 | 4.0 | 1.3 | 1.5 | 2.7 |
| 3e | -0.7 | -1.1 | -2.2 | -2.7 |
| 3a | -2.7 | -2.2 | -1.2 | -2.2 |
| 8 | -6.2 | -5.8 | -4.0 | -4.0 |
| 9 | -0.5 | -0.2 | -0.4 | 0.8 |
| 10 | -1.2 | -0.4 | 0.6 | 0.4 |

### Chiral polymers

**Table A4-51. Comparison of extracted RDCs ($^1D_{CH}$) in Hz of menthol aligned in the L/D-TEG-Phe/DMAA (1,2,3 = (+) and 3,4 = (-))**

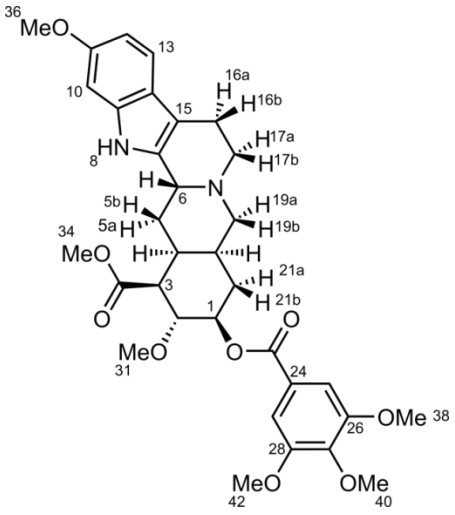
| H | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | -8.5 | -9.5 | -9.4 | -8.8 | -10.7 |
| 2 | -11.4 | -11.4 | -11.4 | -11.4 | -11.4 |
| 3a | -7.0 | -8.0 | -10.0 | -5.6 | -5.2 |
| 3e | 1.4 | 0.8 | 0.6 | -0.9 | -2.1 |
| 4a | -8.3 | -7.2 | -9.7 | -7.3 | -8.0 |
| 4e | 1.5 | 0.5 | 1.5 | -1.9 | -1.4 |
| 5 | -6.4 | -5.8 | -6.5 | -3.9 | -6.0 |
| 6a | -7.3 | -7.3 | -9.1 | -6.3 | -7.2 |
| 6e | 0.0 | 0.9 | 0.9 | -1.0 | -1.8 |
| 7 | -2.0 | -1.9 | -1.9 | -1.8 | -2.0 |
| 8 | -0.1 | -1.4 | -1.2 | -0.9 | -2.5 |
| 9 | 1.6 | 1.4 | 1.7 | 0.9 | 1.4 |
| 10 | 0.1 | -0.2 | 0.8 | 0.4 | 0.8 |



## A4.4.3. Reserpine

**Table A4-52. NMR assignment of reserpine**

| # | $\delta_H$ [ppm] (int,mult,*J* [Hz]) | $\delta_C$ [ppm] |
|---|---|---|
| 1 | 4.99 (1H, ddd, 11.6,9.2,5.1) | 77.7 |
| 2 | 3.85 (1H, m) | 77.8 |
| 3 | 2.64 (1H, dd, 11.2,4.9) | 51.7 |
| 4 | 1.99 (1H, dt, 12.9,5.3) | 32.0 |
| 5a | 2.28 (1H, m) | 24.3 |
| 5b | 1.77 (1H, m) | 24.3 |
| 6 | 4.43 (1H, br. s) | 53.7 |
| 7 | | 129.9 |
| 8 | 7.58 (1H, s) | |
| 9 | | 136.2 |
| 10 | 6.78 (1H, d, 1.9) | 95.2 |
| 11 | | 156.2 |
| 12 | 6.73 (1H, dd, 6.8,2.0) | 109.1 |
| 13 | 7.27 (1H, m) | 118.6 |
| 14 | | 122.1 |
| 15 | | 108 |
| 16a | 2.89 (1H, m) | 16.6 |
| 16b | 2.45 (1H, m) | 16.6 |
| 17a | 3.12 (1H, m) | 51.1 |
| 17b | 3.12 (1H, m) | 51.1 |
| 18 | | |
| 19a | 2.95 (1H, m) | 48.9 |
| 19b | 2.41 (1H, m) | 48.9 |
| 20 | 1.85 (1H, br. d, 13.1) | 33.8 |
| 21a | 2.31 (1H, m) | 29.6 |
| 21b | 1.93 (1H, dt, 13.1,4.2) | 29.6 |
| 22 | | |
| 23 | | 165.4 |
| 24 | | 125.3 |
| 25 | 7.26 (1H, s) | 106.8 |
| 26 | | 152.8 |
| 27 | | 142.1 |
| 28 | | 152.8 |
| 29 | 7.26 (1H, s) | 106.8 |
| 30 | | |
| 31 | 3.44 (3H,s) | 60.7 |
| 32 | | 172.7 |
| 33 | | |
| 34 | 3.76 (3H,s) | 51.8 |
| 35 | | |
| 36 | 3.78 (3H,s) | 55.7 |
| 37 | | |
| 38 | 3.85 (3H,s) | 56.2 |
| 39 | | |
| 40 | 3.85 (3H,s) | 60.8 |
| 41 | | |
| 42 | 3.85 (3H,s) | 56.2 |



## RDC

**Table A4-53. Experimental distances from relative NOE intensities in Å and *J*-coupling constants in Hz compared to best fit of multiple structures for reserpine ($\tau_m$ = 150 ms)**

| H1 | H2 | $\eta^{norm}$ | Exp | Calc | Diff | % |
|---|---|---|---|---|---|---|
| 1 | 3 | 0.00693 | 2.62 | 2.70 | -0.08 | 2.9 |
| 1 | 31 | 0.00211 | 3.83 | 3.74 | 0.09 | 2.5 |
| 1 | 20 | 0.00695 | 2.62 | 2.66 | -0.04 | 1.6 |
| 1 | 21a | 0.01079 | 2.43 | 2.42 | 0.02 | 0.7 |
| 1 | 21b | 0.00179 | 3.28 | 3.01 | 0.27 | 8.3 |
| 2 | 5b/21b | 0.02338 | 2.14 | 2.28 | -0.14 | 6.4 |
| 2 | 31 | 0.01881 | 2.66 | 2.67 | -0.01 | 0.4 |
| 3 | 4 | 0.01017 | 2.46 | 2.37 | 0.08 | 3.4 |
| 3 | 31 | 0.00356 | 3.53 | 3.63 | -0.11 | 3.0 |
| 4 | 8 | 0.00234 | 3.14 | 3.31 | -0.16 | 5.2 |
| 4 | 19a | 0.00725 | 2.60 | 2.66 | -0.06 | 2.4 |
| 5a | 5b | 0.05216 | 1.87 | 1.75 | 0.12 | 6.6 |
| 5a | 6 | 0.00910 | 2.50 | 2.59 | -0.09 | 3.8 |
| 5a | 8 | 0.00879 | 2.52 | 2.34 | 0.18 | 7.1 |
| 5b | 6 | 0.01457 | 2.31 | 2.33 | -0.02 | 0.7 |
| 6 | 17a/17b | 0.01551 | 2.29 | 2.46 | -0.17 | 7.5 |
| 6 | 19b | 0.00139 | 3.42 | 3.17 | 0.25 | 7.2 |
| 6 | 8 | 0.00135 | 3.44 | 3.17 | 0.27 | 7.8 |
| 10 | 36 | 0.00868 | 3.03 | 2.72 | 0.31 | 10.2 |
| 12 | 36 | 0.00496 | 3.33 | 3.23 | 0.10 | 2.9 |
| 13 | 16b | 0.00217 | 3.18 | 2.94 | 0.24 | 7.6 |
| 16a | 16b | 0.06682 | 1.80 | 1.77 | 0.03 | 1.4 |
| 16a | 19a | 0.00432 | 2.83 | 2.52 | 0.32 | 11.1 |
| **19a** | **19b** | **0.07092** | **1.78** | **1.76** | **0.02** | **1.2** |
| 19a | 20 | 0.01054 | 2.45 | 2.33 | 0.12 | 4.9 |
| 20 | 3 | 0.00864 | 2.52 | 2.58 | -0.06 | 2.5 |
| 20 | 19b | 0.00465 | 2.80 | 2.54 | 0.26 | 9.4 |
| 21a | 19b | 0.00382 | 2.89 | 2.48 | 0.42 | 14.4 |
| 21a | 21b | 0.05533 | 1.85 | 1.74 | 0.11 | 6.0 |

| H1 | H2 | Exp | Calb | Diff |
|---|---|---|---|---|
| 1 | 2 | 9.2 | 8.0 | 1.2 |
| 1 | 21a | 5.2 | 4.9 | 0.3 |
| 1 | 21b | 11.6 | 10.9 | 0.7 |
| 2 | 3 | 11.2 | 11.2 | 0.0 |
| 3 | 4 | 4.9 | 4.1 | 0.8 |
| 4 | 5a | 5.6 | 5.6 | 0.0 |
| 4 | 5b | 12.9 | 11.2 | 1.6 |
| 6 | 5b | 6.7 | 6.5 | 0.2 |
| 6 | 5a | 7.7 | 5.8 | 1.9 |
| 19a | 20 | 4.6 | 6.8 | -2.2 |
| 19b | 20 | 4.7 | 2.6 | 2.1 |
| 20 | 21b | 13.9 | 12.4 | 1.5 |

**Table A4-54. RDC extraction of reserpine (Polymer 8.2 in the thesis). $^{2}$H splitting = 90.4 Hz ($CDCl_3$)**

| **Reserpine** | | | | |
|---|---|---|---|---|
| C | H | *J* [Hz] | *T* [Hz] | *D* [Hz] |
| 1 | 1 | 148.7 | 112.4 | -36.3 |
| 2 | 2 | 143.7 | 108.3 | -35.5 |
| 3 | 3 | 131.1 | 92.6 | -38.5 |
| 4 | 4 | 131.2 | 143.7 | 12.5 |
| 5 | 5a | 127.4 | 108.6 | -18.9 |
| 5 | 5b | 127.5 | 149.9 | 22.5 |
| 6 | 6 | 140.2 | 111.8 | -28.5 |
| 10 | 10 | 158 | 160.3 | 2.3 |
| 12 | 12 | 159 | 186.5 | 27.4 |
| 13 | 13 | 158.8 | 161.1 | 2.4 |
| 16 | 16a | 133.5 | 77.8 | -55.7 |
| 16 | 16b | 129.7 | 166.6 | 36.9 |
| 17 | 17a/17b | 138.6 | 101 | -37.6 |
| 19 | 19a | 132.1 | 147.9 | 15.8 |
| 19 | 19b | 136.8 | 119.2 | -17.6 |
| 20 | 20 | 128.2 | 94.8 | -33.4 |
| 21 | 21a | 132.7 | 105.2 | -27.5 |
| 21 | 21b | 128.1 | 105.2 | -22.9 |
| 25/29 | 25/29 | 163.1 | 146.3 | -16.8 |
| 31 | 31 | 141.9 | 142.3 | 0.4 |
| 34 | 34 | 147 | 154.5 | 7.4 |
| 36 | 36 | 143.4 | 142 | -1.5 |
| 40 | 40 | 145.2 | 143.2 | -2.0 |

**Table A4-55. NOE-distances in Å back-calculated from different assignments of reserpine using ISPA.**

| **NOE** | | | **Right** | **Dia4** | **Dia6** | **Dia19** | **Dia2** | **Dia1** | **Dia21** | **Dia17** | **Dia3** | **Dia5** | **Dia20** | **Dia16** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nuc 1 | Nuc 2 | Exp | Calc | Calc | Calc | Calc | Calc | Calc | Calc | Calc | Calc | Calc | Calc | Calc |
| 8 | 5b | 2.52 | 2.35 | 2.29 | 2.38 | 2.42 | 2.41 | 2.34 | 2.67 | 2.36 | 2.27 | 2.82 | 2.28 | 2.35 |
| 8 | 4 | 3.14 | 3.36 | 4.49 | 4.52 | 3.41 | 3.19 | 4.82 | 2.93 | 3.36 | 3.03 | 3.48 | 3.02 | 3.33 |
| 8 | 34 | 5.00 | 4.91 | 5.48 | 5.39 | 5.14 | 5.05 | 4.62 | 4.77 | 4.92 | 6.36 | 4.87 | 6.86 | 4.84 |
| 8 | 6 | 3.44 | 3.15 | 3.13 | 3.15 | 3.12 | 3.15 | 3.08 | 2.83 | 3.15 | 3.21 | 3.14 | 3.21 | 3.15 |
| 8 | 10 | 3.24 | 2.78 | 2.89 | 2.89 | 2.78 | 2.90 | 2.89 | 2.78 | 2.78 | 2.88 | 2.78 | 2.89 | 2.78 |
| 12 | 36 | 3.33 | 3.23 | 2.68 | 2.68 | 3.23 | 2.59 | 2.59 | 3.06 | 3.23 | 2.70 | 3.19 | 2.59 | 3.23 |
| 1 | 3 | 2.62 | 2.68 | 2.62 | 2.63 | 2.75 | 2.44 | 3.84 | 2.69 | 2.68 | 3.84 | 2.75 | 4.24 | 2.70 |
| 1 | 31 | 3.83 | 3.35 | 3.46 | 3.39 | 3.22 | 4.53 | 4.85 | 3.00 | 3.35 | 3.98 | 3.49 | 2.94 | 3.36 |
| 1 | 20 | 2.62 | 2.62 | 2.48 | 2.62 | 2.62 | 2.47 | 3.81 | 3.29 | 2.62 | 2.62 | 2.66 | 3.83 | 2.65 |
| 1 | 21b | 2.43 | 2.43 | 2.45 | 2.43 | 2.44 | 2.46 | 3.03 | 2.99 | 2.43 | 2.43 | 2.47 | 2.49 | 2.44 |
| 6 | 17a17b | 2.29 | 2.48 | 2.42 | 2.41 | 2.51 | 2.39 | 2.42 | 2.61 | 2.48 | 2.32 | 2.49 | 2.32 | 2.47 |
| 6 | 19b | 3.42 | 3.15 | 3.59 | 3.62 | 3.79 | 2.71 | 2.40 | 3.71 | 3.13 | 4.17 | 2.80 | 4.16 | 3.13 |
| 6 | 5a | 2.31 | 2.33 | 3.06 | 2.83 | 2.34 | 2.33 | 3.05 | 2.92 | 2.33 | 2.35 | 2.67 | 2.34 | 2.34 |
| 6 | 5b | 2.50 | 2.60 | 2.49 | 2.54 | 2.63 | 2.64 | 2.52 | 2.39 | 2.60 | 2.50 | 2.36 | 2.50 | 2.59 |
| 19a | 20 | 2.45 | 2.33 | 3.06 | 2.44 | 2.64 | 2.28 | 2.44 | 2.49 | 2.33 | 2.36 | 2.33 | 3.04 | 2.35 |
| 19a | 4 | 2.60 | 2.70 | 2.60 | 2.58 | 3.79 | 2.78 | 4.24 | 2.47 | 2.70 | 2.57 | 2.81 | 2.57 | 2.71 |
| 17a17b | 19b | 2.02 | 2.38 | 2.23 | 2.23 | 2.49 | 2.34 | 2.46 | 2.26 | 2.38 | 2.40 | 2.38 | 2.40 | 2.40 |
| **19a** | **19b** | **1.78** | **1.76** | **1.76** | **1.75** | **1.75** | **1.75** | **1.75** | **2.61** | **1.76** | **1.75** | **1.76** | **1.75** | **1.76** |
| 16a | 16b | 1.79 | 1.77 | 1.75 | 1.75 | 1.77 | 1.74 | 1.75 | 2.66 | 1.77 | 1.74 | 1.77 | 1.74 | 1.77 |
| 3 | 4 | 2.46 | 2.40 | 3.05 | 2.39 | 2.39 | 2.36 | 2.29 | 2.29 | 2.40 | 2.54 | 2.38 | 2.36 | 2.38 |
| 21a | 21b | 1.85 | 1.74 | 1.75 | 1.74 | 1.74 | 1.75 | 1.74 | 1.76 | 1.74 | 1.74 | 1.74 | 1.74 | 1.74 |
| 5a | 5b | 1.87 | 1.75 | 1.75 | 1.74 | 1.75 | 1.73 | 1.74 | 2.04 | 1.75 | 1.74 | 1.75 | 1.74 | 1.75 |
| 20 | 3 | 2.52 | 2.58 | 2.59 | 2.52 | 2.63 | 2.52 | 4.22 | 1.75 | 2.58 | 3.82 | 2.67 | 3.85 | 2.61 |
| 21b | 19b | 2.89 | 2.49 | 2.51 | 2.53 | 3.19 | 2.48 | 3.92 | 2.63 | 2.49 | 2.53 | 2.55 | 2.54 | 2.52 |
| 20 | 19b | 2.80 | 2.56 | 2.46 | 2.48 | 2.32 | 2.62 | 3.05 | 1.74 | 2.56 | 2.48 | 2.63 | 2.45 | 2.57 |
| 5a21a | 2 | 2.14 | 2.26 | 2.72 | 2.31 | 2.28 | 3.48 | 4.18 | 1.76 | 2.25 | 2.16 | 2.71 | 3.73 | 2.27 |
| 3 | 31 | 3.51 | 3.40 | 3.39 | 3.42 | 3.28 | 4.73 | 2.83 | 2.87 | 3.40 | 3.57 | 3.46 | 3.19 | 3.46 |
| 21a | 6 | 2.36 | 2.72 | 4.32 | 4.19 | 2.62 | 2.90 | 5.09 | 2.56 | 2.70 | 3.78 | 2.83 | 5.53 | 2.85 |
| 21a | 1 | 3.28 | 3.05 | 3.04 | 3.04 | 3.06 | 3.04 | 2.46 | 4.35 | 3.05 | 3.05 | 3.05 | 2.36 | 3.03 |
| 16b | 13 | 3.18 | 2.93 | 2.93 | 3.34 | 2.91 | 2.96 | 2.96 | 2.75 | 2.93 | 3.00 | 2.92 | 3.00 | 3.26 |
| 31 | 2 | 2.66 | 2.74 | 2.75 | 2.76 | 2.78 | 2.66 | 2.72 | 3.71 | 2.74 | 2.72 | 2.71 | 2.72 | 2.73 |
| 10 | 36 | 3.03 | 2.72 | 3.34 | 3.34 | 2.72 | 4.50 | 4.49 | 2.40 | 2.72 | 3.25 | 2.73 | 4.49 | 2.72 |
| 16a | 19a | 2.83 | 2.55 | 4.70 | 4.83 | 3.51 | 2.56 | 4.15 | 3.47 | 2.55 | 2.32 | 2.66 | 2.31 | 3.96 |

**Table A4-56. *J*-coupling constants in Hz back-calculated from different assignments of reserpine using HLA.**

| **J-coupl** | | | **Right** | **Dia4** | **Dia6** | **Dia19** | **Dia2** | **Dia1** | **Dia21** | **Dia17** | **Dia3** | **Dia5** | **Dia20** | **Dia16** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nuc 1 | Nuc 2 | Exp | Calc | Calc | Calc | Calc | Calc | Calc | Calc | Calc | Calc | Calc | Calc | Calc |
| 1 | 2 | 9.2 | 7.9 | 8.9 | 8.4 | 7.6 | 2.6 | 2.4 | 7.4 | 7.9 | 9 | 7.5 | 2.3 | 7.7 |
| 1 | 21b | 5.2 | 4.8 | 4 | 4.9 | 4.7 | 3.9 | 4.2 | 9.2 | 4.8 | 4.7 | 4.5 | 1.6 | 4.9 |
| 1 | 21a | 11.6 | 11 | 11.6 | 11 | 11 | 11.6 | 11.4 | 5.1 | 11 | 11.2 | 10.9 | 3.4 | 10.5 |
| 2 | 3 | 11.2 | 11 | 12.4 | 12.5 | 9.7 | 2 | 0.7 | 10.8 | 11 | 5.3 | 9.3 | 0.5 | 10.9 |
| 3 | 4 | 4.9 | 4.1 | 14.2 | 3.6 | 4.1 | 4 | 5.6 | 5 | 4.1 | 1.2 | 4.1 | 4.7 | 4.4 |
| 4 | 5b | 5.6 | 4.6 | 3.3 | 5.6 | 5.3 | 4.7 | 6.4 | 4.8 | 4.6 | 2.9 | 9.4 | 2.8 | 4.5 |
| 4 | 5a | 12.9 | 11.8 | 12.5 | 10.3 | 11.3 | 11.5 | 1.3 | 10.2 | 11.8 | 12.6 | 5.3 | 12.6 | 11.4 |
| 6 | 5a | 6.7 | 6 | 12 | 9.8 | 5.9 | 5.8 | 2.9 | 5.4 | 6 | 5.3 | 5.9 | 5.1 | 5.6 |
| 6 | 5b | 7.7 | 4.1 | 3.2 | 5 | 5 | 5.6 | 12 | 5.2 | 4.2 | 2.6 | 5.3 | 2.7 | 4.1 |
| 19a | 20 | 4.6 | 5.3 | 3.5 | 3.4 | 4 | 6.8 | 3.3 | 5.1 | 5.4 | 3.9 | 5.4 | 12 | 5 |
| 19b | 20 | 4.7 | 3.1 | 12 | 3.2 | 5.8 | 2.7 | 12 | 4.4 | 3.1 | 2.9 | 4.7 | 4 | 3.8 |
| 20 | 21a | 13.9 | 12.4 | 12.6 | 12.5 | 12.2 | 12.6 | 5.4 | 3.7 | 12.4 | 12.4 | 11.7 | 12.5 | 11.9 |

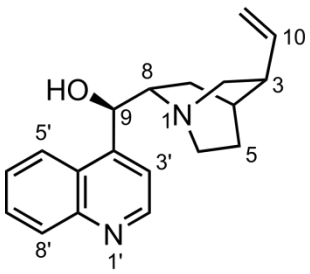
**Table A4-57. RDCs in Hz back-calculated from different assignments of reserpine using the *θ*-method (Polymer 8.2 in the thesis).**

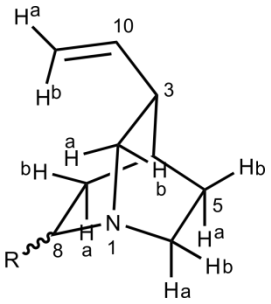
| **RDC** | | | **Right** | **Dia4** | **Dia6** | **Dia19** | **Dia2** | **Dia1** | **Dia21** | **Dia17** | **Dia3** | **Dia5** | **Dia20** | **Dia16** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C | H | Exp | Calc | Calc | Calc | Calc | Calc | Calc | Calc | Calc | Calc | Calc | Calc | Calc |
| 13 | 13 | 2.4 | 1.7 | 7.3 | 9.6 | 0.7 | 2.9 | -0.2 | 2.8 | 2.0 | 3.1 | 2.3 | 6.5 | 4.1 |
| 12 | 12 | 27.4 | 29.4 | 16.6 | 15.5 | 28.8 | 24.9 | 31.9 | 28.8 | 29.2 | 24.9 | 27.6 | 11.0 | 31.9 |
| 25/29 | 25/29 | -16.8 | -18.8 | -18.6 | -26.9 | -17.7 | 1.3 | -17.3 | -17.2 | -18.9 | -17.6 | -17.4 | -11.8 | -13.2 |
| 10 | 10 | 2.3 | 2.2 | 7.2 | 9.8 | 0.8 | 3.0 | -0.2 | 2.9 | 2.4 | 3.3 | 2.6 | 6.5 | 4.7 |
| 2 | 2 | -35.5 | -39.5 | -26.6 | -27.6 | -39.0 | 11.8 | -24.2 | -38.2 | -39.5 | -35.9 | -34.5 | 27.6 | -27.3 |
| 1 | 1 | -36.3 | -36.4 | -25.8 | -30.7 | -32.3 | -31.2 | -10.2 | -36.8 | -36.3 | -35.2 | -31.8 | -17.6 | -26.0 |
| 40 | 40 | -2.0 | -2.4 | -3.3 | -4.5 | -5.1 | 0.6 | -2.0 | -1.9 | -2.5 | -2.1 | -5.2 | -4.3 | -2.6 |
| 31 | 31 | 0.4 | 0.3 | -0.9 | -2.7 | 2.3 | -5.0 | 1.9 | 1.6 | 0.1 | 0.8 | 2.7 | -3.4 | -4.1 |
| 36 | 36 | -1.5 | -2.0 | -3.8 | -6.5 | -1.9 | -8.4 | -10.6 | -1.9 | -1.9 | -4.9 | -1.6 | -3.4 | -1.2 |
| 6 | 6 | -28.5 | -30.9 | -19.9 | -5.2 | -31.2 | -26.5 | -33.7 | -28.7 | -30.6 | -31.3 | -29.7 | -16.0 | -12.0 |
| 34 | 34 | 7.4 | 5.8 | 7.7 | 5.4 | 2.0 | 0.1 | 4.0 | 6.6 | 5.9 | 5.5 | 1.7 | -0.8 | 1.5 |
| 3 | 3 | -38.5 | -34.9 | -20.7 | -25.2 | -44.3 | -32.1 | -7.4 | -36.4 | -34.4 | -37.0 | -37.6 | -14.3 | -26.0 |
| 17 | 17a/17b | -37.6 | -29.5 | -22.5 | -14.3 | -27.0 | -22.0 | -20.1 | -30.2 | -29.2 | -27.0 | -28.9 | -18.0 | -6.9 |
| 19 | 19a | 15.8 | 12.5 | -19.3 | -2.8 | 5.8 | 21.5 | -2.9 | 12.7 | 12.4 | 16.8 | -2.7 | 6.5 | 21.3 |
| 19 | 19b | -17.6 | -20.5 | -21.8 | -20.3 | -11.6 | -24.5 | -34.3 | -20.0 | -20.0 | -20.7 | -11.7 | -16.8 | -4.0 |
| 20 | 20 | -33.4 | -32.5 | -25.6 | -22.7 | -29.5 | -30.2 | -31.7 | -33.7 | -32.5 | -35.6 | -35.8 | 7.5 | -27.6 |
| 4 | 4 | 12.5 | 16.9 | -24.3 | -3.7 | 18.4 | 17.6 | -11.2 | 15.6 | 17.2 | 19.0 | 12.0 | 7.9 | 23.7 |
| 21 | 21b | -22.9 | -24.8 | -21.3 | -18.3 | -24.6 | -28.3 | -3.8 | -23.5 | -24.6 | -24.5 | -25.8 | 7.6 | -22.0 |
| 21 | 21a | -27.5 | -25.3 | -18.0 | -22.0 | -22.9 | -9.3 | -7.9 | -27.0 | -25.3 | -30.8 | -23.7 | -19.4 | -17.4 |
| 5 | 5b | 22.5 | 19.3 | -17.4 | 7.7 | 15.6 | 20.5 | -4.0 | 19.6 | 19.2 | 16.2 | 3.8 | 7.2 | 27.3 |
| 5 | 5a | -18.9 | -23.3 | -23.9 | -33.3 | -17.2 | -11.0 | -31.9 | -21.0 | -22.8 | -23.7 | -6.4 | -20.5 | 1.2 |
| 16 | 16a | -55.7 | -55.7 | -44.1 | -14.5 | -48.9 | -46.7 | -31.0 | -56.3 | -55.2 | -54.1 | -55.3 | -38.4 | -22.0 |
| 16 | 16b | 36.9 | 33.6 | 27.2 | 43.1 | 38.7 | 7.6 | 9.8 | 33.2 | 33.9 | 34.3 | 30.2 | 29.3 | 4.8 |

## A4.4.4. Cinchonidine

**Table A4-58. NMR assignment of cinchonidine**

| # | $\delta_H$ [ppm] (int,mult,*J* [Hz]) | $\delta_C$ [ppm] |
|---|---|---|
| 2' | 8.80 (1H,d,4.5) | 150.1 |
| 3' | 7.61 (1H,m) | 118.3 |
| 4' | | 149.1 |
| 5' | 7.95 (1H,d,8,4) | 122.9 |
| 6' | 7.60 (1H,m) | 129.0 |
| 7' | 7.30 (1H,t,8.0) | 126.6 |
| 8' | 8.05 (1H,d,8.4) | 12.30 |
| 9' | | 147.7 |
| 10' | | 125.6 |
| 2a | 3.11 (1H,dd,13.4,10.2) | 56.5 |
| 2b | 2.71 (1H,m) | 56.5 |
| 3 | 2.32 (1H,br. t,9.8) | 39.3 |
| 4 | 1.85 (1H,m) | 27.6 |
| 5a | 1.83 (1H,m) | 27.0 |
| 5b | 1.54 (1H,m) | 27.0 |
| 6a | 3.66 (1H,td,10.1,3.3) | 43.2 |
| 6b | 2.68 (1H,m) | 43.2 |
| 7a | 1.83 (1H,m) | 20.9 |
| 7b | 1.52 (1H,m) | 20.9 |
| 8 | 3.13 (1H,m) | 60.2 |
| 9 | 5.81 (1H,d,3.0) | 70.7 |
| 10 | 5.68 (1H,ddd,17.3,10.1,7.3) | 141.0 |
| 11a | 4.95 (1H,m) | 114.8 |
| 11b | 4.92 (1H,m) | 114.8 |
| OH | - | - |




## RDC

**Table A4-59. RDC extraction of cinchonidine (Polymer 8.3 in the thesis). $^2$H splitting = 67.4 Hz ($CDCl_3$)**

| **Cinchonidine** | | | | |
|---|---|---|---|---|
| C | H | *J* [Hz] | *T* [Hz] | *D* [Hz] |
| 2 | 2a | 139.6 | 143.5 | 3.9 |
| 2 | 2b | 138.5 | 149.2 | 10.8 |
| 3 | 3 | 130.1 | 154.5 | 24.5 |
| 4 | 4 | 139.1 | 167.1 | 27.9 |
| 5 | 5a | 131 | 112.4 | -18.5 |
| 5 | 5b | 128.1 | 157.8 | 29.7 |
| 6 | 6a | 141.6 | 103 | -38.6 |
| 6 | 6b | 137.8 | 142.7 | 5 |
| 7 | 7a | 131 | 110 | -21 |
| 7 | 7b | 129.8 | 107 | -22.7 |
| 8 | 8 | 136.4 | 131.1 | -5.3 |
| 9 | 9 | 142.9 | 158.7 | 15.8 |
| 10 | 10 | 151 | 177.2 | 26.2 |
| 11 | 11a | 158.4 | 131.4 | -27 |
| 11 | 11b | 153.4 | 174.7 | 21.3 |
| 2' | 2' | 177.7 | 200 | 22.3 |
| 3' | 3' | 163.7 | 187.2 | 23.5 |
| 5' | 5' | 159.4 | 187.2 | 27.8 |
| 6' | 6' | 160.4 | 185.5 | 25.1 |
| 7' | 7' | 160.8 | 181.7 | 20.9 |
| 8' | 8' | 162.2 | 187.8 | 25.6 |

**Table A4-60. RDCs in Hz back-calculated from different assignments of cinchonidine using the *θ*-method (Polymer 8.3 in the thesis).**

| C | H | Epx | Right | Dia-9 | Dia-7 | Dia-2 | Dia-6 | Dia-5 | Dia-11 |
|---|---|---|---|---|---|---|---|---|---|
| 2' | 2' | 22.3 | 24.3 | 20.1 | 23.9 | 24.6 | 24.3 | 23.8 | 23.9 |
| 10 | 10 | 26.2 | 23.9 | 22.7 | 24.0 | 23.8 | 21.2 | 23.1 | 7.1 |
| 8' | 8' | 25.6 | 25.1 | 18.4 | 25.3 | 25.3 | 24.2 | 25.3 | 23.4 |
| 6' | 6' | 25.1 | 23.0 | 20.7 | 22.6 | 23.3 | 23.2 | 23.3 | 23.3 |
| 7' | 7' | 20.9 | 21.5 | 15.4 | 21.4 | 21.1 | 19.8 | 22.6 | 20.6 |
| 5' | 5' | 27.8 | 25.2 | 18.0 | 25.4 | 25.4 | 24.1 | 24.9 | 23.9 |
| 3' | 3' | 23.5 | 22.3 | 15.4 | 22.2 | 22.0 | 20.6 | 22.1 | 21.0 |
| 11 | 11b | 21.3 | 24.4 | 24.1 | 24.6 | 24.6 | 25.9 | 24.8 | -1.3 |
| 11 | 11a | -27.0 | -28.8 | -26.7 | -28.2 | -28.1 | -24.0 | -27.8 | 2.5 |
| 9 | 9 | 15.8 | 16.1 | -1.7 | 16.2 | 16.7 | 16.7 | 15.2 | 16.2 |
| 8 | 8 | -5.3 | -7.0 | -2.2 | -7.3 | -5.3 | -10.9 | -12.2 | -4.9 |
| 2 | 2a | 3.9 | 4.7 | 13.0 | 4.9 | 6.5 | 3.9 | 11.2 | 4.9 |
| 2 | 2b | 10.8 | 8.3 | 15.9 | 8.5 | 8.2 | -13.4 | 7.9 | 6.1 |
| 6 | 6a | -38.6 | -37.7 | -41.3 | -37.9 | -37.1 | -14.4 | -35.2 | -27.7 |
| 6 | 6b | 5.0 | 9.0 | 14.8 | 9.1 | 6.7 | -2.0 | 9.4 | 6.8 |
| 3 | 3 | 24.5 | 23.2 | 25.3 | 23.3 | 23.0 | 20.9 | 3.1 | 18.3 |
| 4 | 4 | 27.9 | 23.1 | 12.5 | 23.2 | 22.4 | 17.5 | 21.2 | 14.6 |
| 5 | 5a | -18.5 | -21.2 | -12.9 | -21.9 | -20.6 | -19.0 | 1.5 | -24.9 |
| 5 | 5b | 29.7 | 25.8 | 23.4 | 25.8 | 24.4 | 23.0 | 7.3 | 17.1 |
| 7 | 7a | -22.7 | -22.3 | -15.7 | -21.5 | -21.2 | -23.3 | -19.7 | -4.3 |
| 7 | 7b | -21.0 | -20.3 | -17.3 | -21.3 | -20.0 | -19.3 | -5.6 | -24.5 |
| | | *Q*-factor | 0.100 | 0.331 | 0.102 | 0.106 | 0.372 | 0.410 | 0.497 |

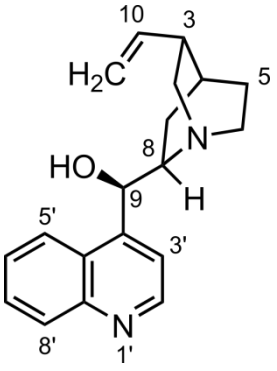
**Table A4-61. RDC extraction of cinchonidine (Polymer 8.4 in the thesis). $^2$H splitting = N/A (DMSO-$d_6$)**

| **Cinchonidine** | | | | |
|---|---|---|---|---|
| C | H | *J* [Hz] | *T* [Hz] | *D* [Hz] |
| 2' | 2' | 177.6 | 190.1 | 12.5 |
| 10 | 10 | 150.8 | 162.7 | 11.9 |
| 8' | 8' | 161.6 | 175.8 | 14.2 |
| 6' | 6' | 160.1 | 173.3 | 13.2 |
| 7' | 7' | 160.4 | 170.0 | 9.6 |
| 5' | 5' | 160.7 | 176.3 | 15.6 |
| 3' | 3' | 162.7 | 173.5 | 10.8 |
| 11 | 11b | 153.8 | 157.8 | 4.0 |
| 9 | 9 | 141.4 | 140.4 | -1.0 |
| 8 | 8 | 135.6 | 136.9 | 1.4 |
| 2 | 2b | 138.5 | 139.5 | 1.1 |
| 2 | 2a | 136.7 | 138.3 | 1.6 |
| 6 | 6a | 138.6 | 132.0 | -6.6 |
| 6 | 6b | 137.5 | 136.5 | -1.1 |
| 3 | 3 | 129.8 | 133.8 | 4.1 |
| 4 | 4 | 138.1 | 147.2 | 9.2 |
| 5 | 5a | 130.8 | 137.7 | 6.9 |
| 5 | 5b | 129.1 | 128.8 | -0.3 |

## A4.4.5. Cinchonine

**Table A4-62. NMR assignment of cinchonine**

| # | $\delta_H$ [ppm] (int,mult,*J* [Hz]) | $\delta_C$ [ppm] |
|---|---|---|
| 2' | 8.90 (1H,d,4.4) | 150.2 |
| 3' | 7.64 (1H,m) | 118.2 |
| 4' | | 148.4 |
| 5' | 7.98 (1H,d,8.4) | 122.9 |
| 6' | 7.45 (1H,d,8.4) | 126.8 |
| 7' | 7.68 (1H,t,7.2) | 129.0 |
| 8' | 8.12 (1H,d,8.2) | 130.3 |
| 9' | | 148.0 |
| 10' | | 125.3 |
| 2a | 2.97 (1H,m) | 49.3 |
| 2b | 3.41 (1H,ddd,12.8,8.3,1.2) | 49.3 |
| 3 | 2.28 (1H,q,8.3) | 39.6 |
| 4 | 1.80 (1H,br. s) | 28.0 |
| 5a | 1.57 (1H,m) | 25.9 |
| 5b | 1.54 (1H,m) | 25.9 |
| 6a | 2.83 (1H,m) | 50.0 |
| 6b | 2.94 (1H,m) | 50.0 |
| 7a | 2.06 (1H,dd,12.1,10.4) | 20.9 |
| 7b | 1.18 (1H, dqd, 13.5,4.6,1.1) | 20.9 |
| 8 | 3.14 (1H,td,8.9,3.9) | 60.0 |
| 9 | 5.84 (1H,d,4.0) | 71.2 |
| 10 | 6.03 (1H,ddd,17.8,9.7,7.7) | 140.0 |
| 11a | 5.02 (1H,m) | 114.8 |
| 11b | 5.01 (1H,m) | 114.8 |
| OH | 2.15 (1H,br. s) | - |




## RDC

**Table A4-63. RDC extraction of cinchonine (Polymer 8.5 in the thesis). $^2$H splitting = 56.5 Hz ($CDCl_3$)**

| **Cinchonine** | | | | |
|---|---|---|---|---|
| C | H | *J* [Hz] | *T* [Hz] | *D* [Hz] |
| 2 | 2a | 140.6 | 107.5 | -33.1 |
| 2 | 2b | 138.9 | 147.9 | 9 |
| 3 | 3 | 129 | 127.5 | -1.5 |
| 4 | 4 | 139.8 | 143.5 | 3.7 |
| 5 | 5a | 130.8 | 125.5 | -5.3 |
| 5 | 5b | 128.3 | 124.7 | -3.6 |
| 6 | 6a | 137.6 | 156.7 | 19.2 |
| 6 | 6b | 137.5 | 145.5 | 8.1 |
| 7 | 7a | 130.9 | 120.3 | -10.6 |
| 7 | 7b | 128.8 | 111.8 | -17 |
| 8 | 8 | 135.7 | 139 | 3.2 |
| 9 | 9 | 143.7 | 157.6 | 13.8 |
| 10 | 10 | 151.9 | 144.4 | -7.6 |
| 2' | 2' | 177.7 | 207.8 | 30.1 |
| 3' | 3' | 163.4 | 188.1 | 24.7 |
| 5' | 5' | 159.2 | 183.1 | 23.9 |
| 6' | 6' | 160.3 | 189.8 | 29.5 |
| 7' | 7' | 167.3 | 190.3 | 23 |
| 8' | 8' | 162.3 | 187 | 24.8 |

**Table A4-64. RDCs in Hz back-calculated from different assignments of cinchonine using the *θ*-method (Polymer 8.5 in the thesis).**

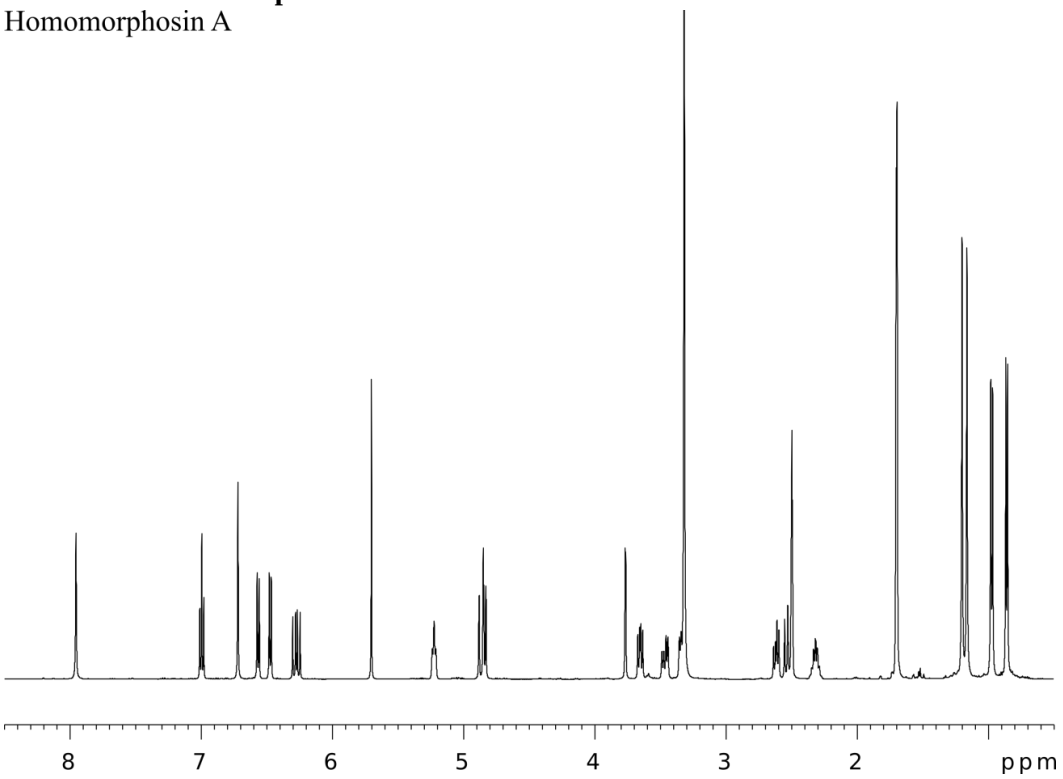
| C | H | Exp | Right | Dia-9 | Dia-7 | Dia-2 | Dia-6 | Dia-5 | Dia-11 | Right-30kJ/mol | Dia-9-30kJ/mol |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2' | 2' | 30.06 | 29.8 | 29.9 | 29.9 | 29.5 | 30.1 | 29.9 | 30.0 | 29.9 | 27.7 |
| 10 | 10 | -7.57 | -7.3 | -7.6 | -7.3 | -7.4 | -7.8 | -7.3 | -7.5 | -7.5 | -7.6 |
| 8' | 8' | 24.76 | 24.2 | 24.0 | 24.2 | 23.8 | 24.1 | 24.1 | 24.0 | 24.1 | 17.5 |
| 7' | 7' | 22.97 | 23.4 | 23.0 | 23.1 | 22.6 | 22.7 | 23.1 | 22.9 | 23.0 | 19.0 |
| 6' | 6' | 29.45 | 29.5 | 29.3 | 29.5 | 29.2 | 29.7 | 29.6 | 29.8 | 29.6 | 28.4 |
| 5' | 5' | 23.89 | 24.6 | 24.7 | 24.7 | 23.9 | 24.7 | 24.5 | 24.5 | 24.4 | 17.6 |
| 3' | 3' | 24.68 | 24.8 | 24.6 | 24.4 | 24.1 | 24.4 | 24.7 | 24.4 | 24.3 | 20.7 |
| 9 | 9 | 13.83 | 13.9 | 13.8 | 13.6 | 13.3 | 13.6 | 13.7 | 13.5 | 13.6 | 6.8 |
| 8 | 8 | 3.23 | 0.9 | 1.9 | 1.3 | 2.7 | -1.6 | 1.1 | 0.6 | 0.5 | 7.2 |
| 6 | 6a | 19.15 | 21.1 | 20.2 | 21.5 | 15.6 | 14.2 | 21.3 | 20.8 | 21.1 | 21.5 |
| 6 | 6b | 8.06 | 8.6 | 8.8 | 9.1 | -12.2 | 13.1 | 8.6 | 9.3 | 9.3 | 14.7 |
| 2 | 2a | -33.14 | -33.4 | -33.3 | -33.4 | -19.0 | -33.0 | -33.0 | -34.0 | -33.9 | -36.6 |
| 2 | 2b | 9.03 | 7.6 | 8.1 | 8.0 | 0.4 | 13.5 | 7.9 | 8.3 | 7.9 | 13.8 |
| 3 | 3 | -1.5 | -0.7 | -0.3 | -0.6 | 0.7 | 1.5 | -1.3 | -0.6 | -0.6 | 1.6 |
| 4 | 4 | 3.67 | 6.9 | 7.5 | 6.1 | 2.7 | 7.3 | 6.9 | 6.8 | 6.9 | 2.1 |
| 5 | 5a | -5.29 | -3.0 | -3.4 | -6.2 | -1.7 | -3.0 | -2.5 | -2.9 | -3.1 | -1.3 |
| 5 | 5b | -3.59 | -1.2 | -1.5 | -1.4 | 5.8 | 0.2 | -2.2 | -1.2 | -1.4 | -0.2 |
| 7 | 7a | -16.98 | -13.1 | -12.9 | -12.0 | -8.3 | -12.1 | -12.9 | -13.1 | -12.8 | -9.2 |
| 7 | 7b | -10.63 | -7.5 | -7.3 | -7.2 | -7.9 | -6.4 | -6.8 | -7.4 | -7.9 | -6.1 |
| | | *Q*-factor | 0.096 | 0.095 | 0.098 | 0.373 | 0.167 | 0.099 | 0.098 | 0.097 | 0.250 |

# A5. NMR spectra

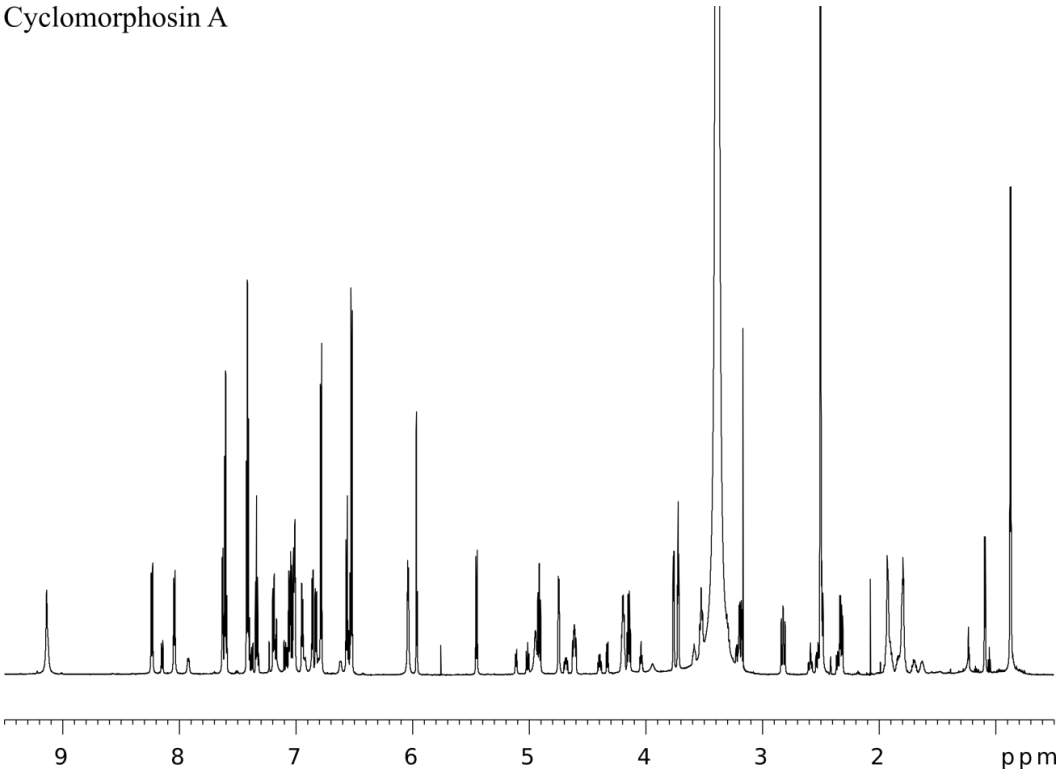
## A5.1. 1D spectra

### A5.1.1. Natural compounds

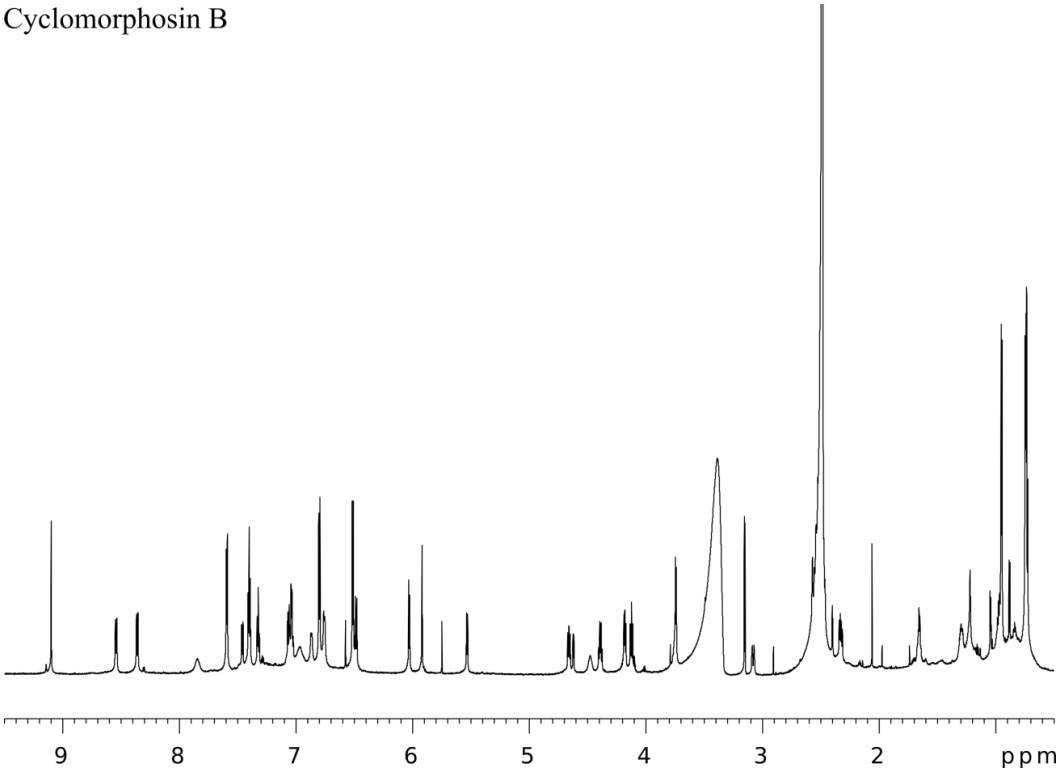
Homomorphosin A



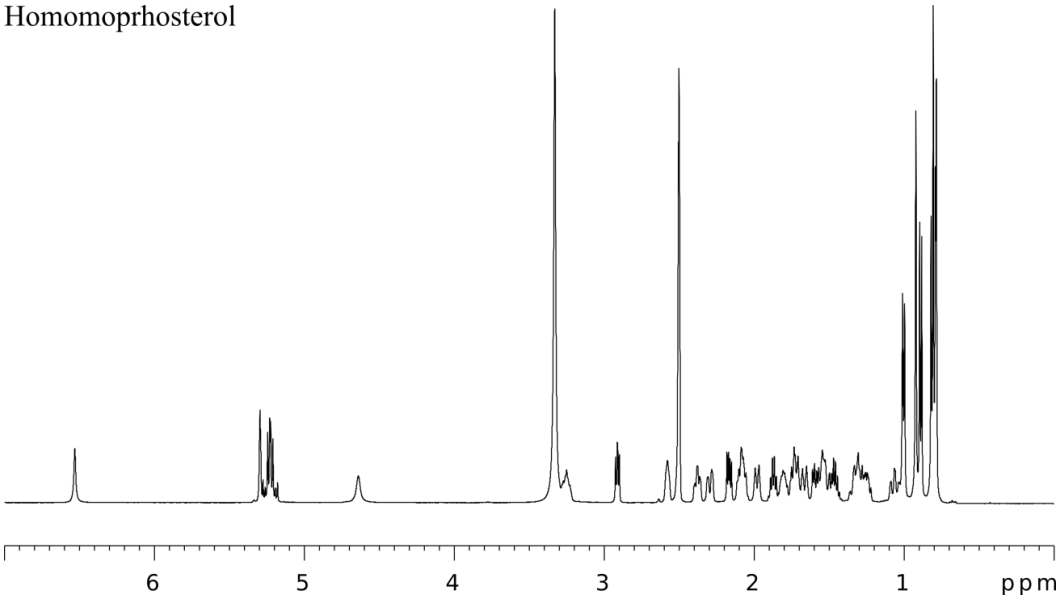

Cyclomorphosin A



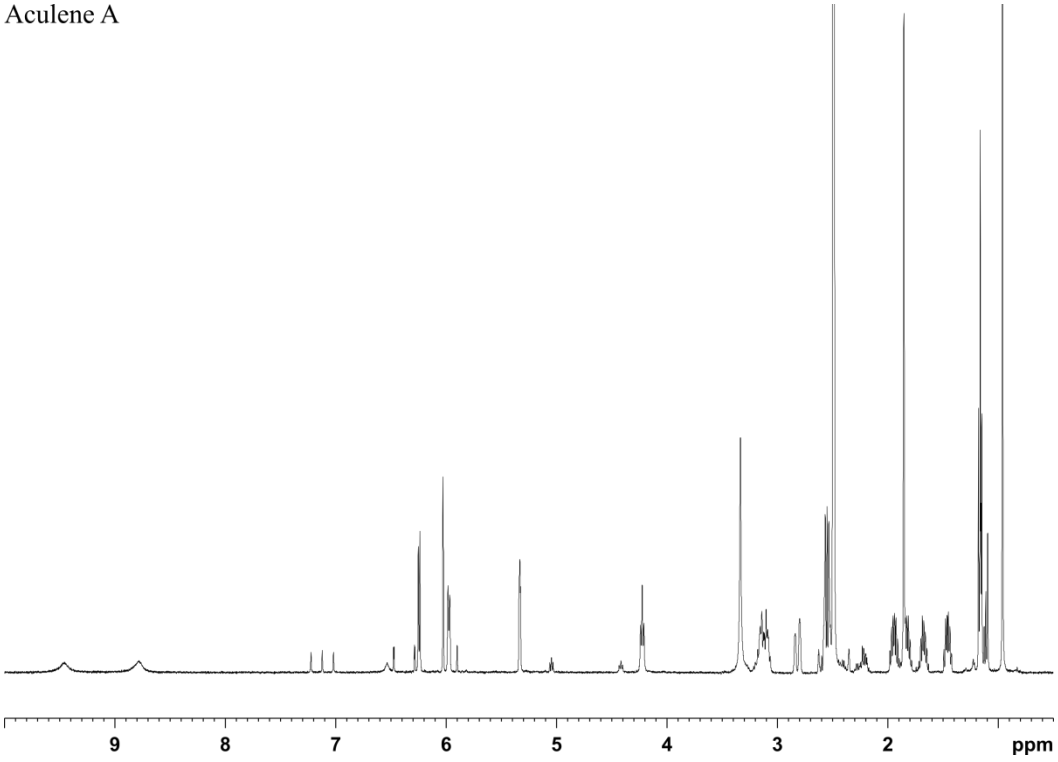
Cyclomorphosin B



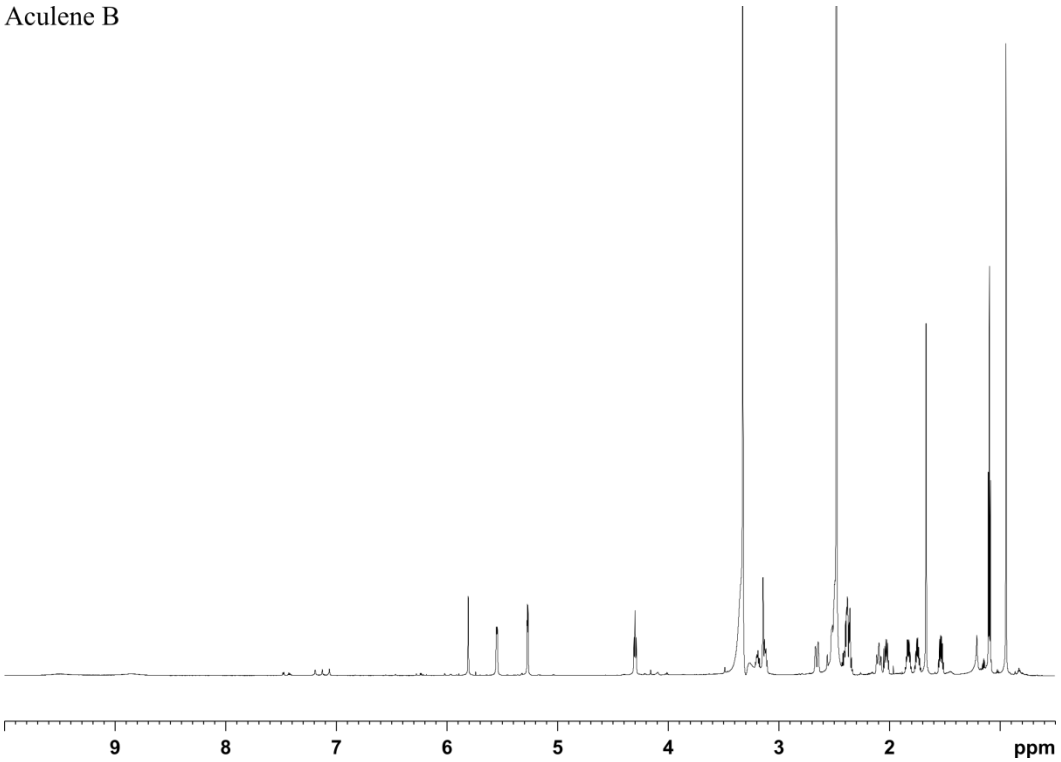

Homomoprhosterol



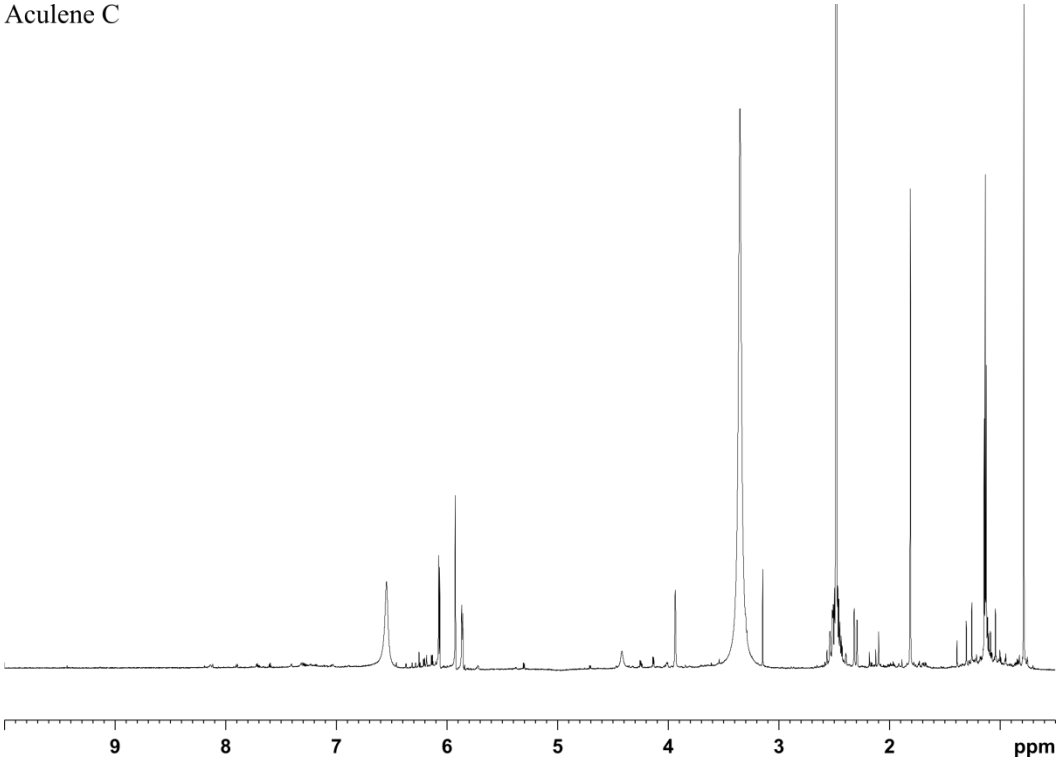
Aculene A



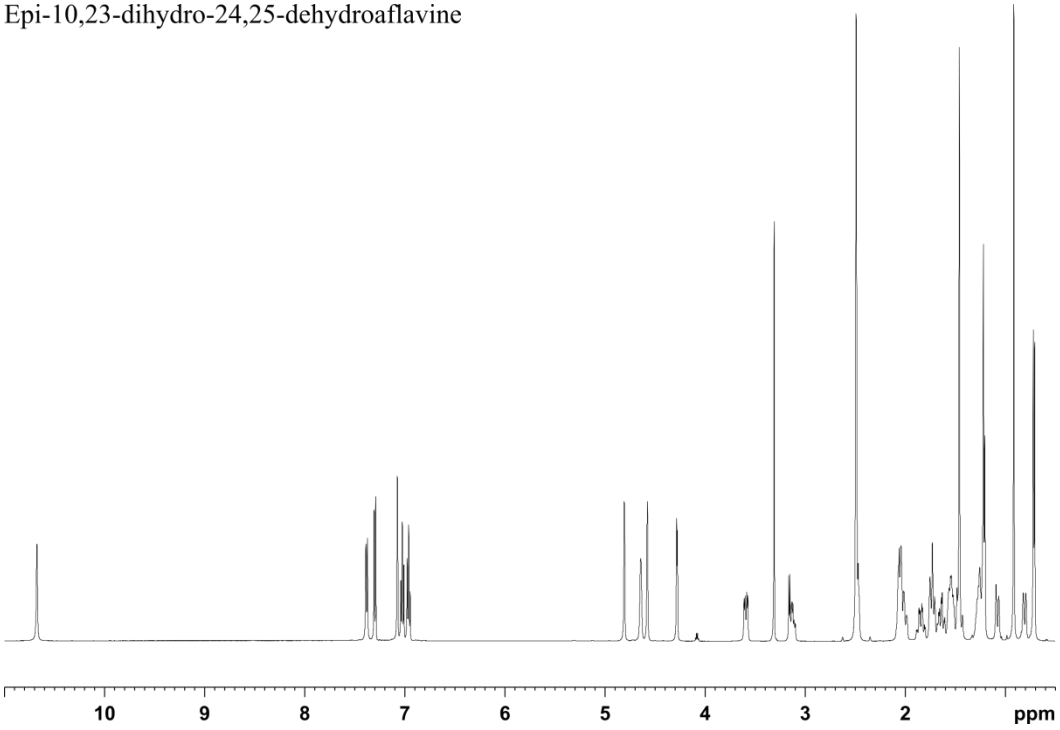

Aculene B



Aculene C

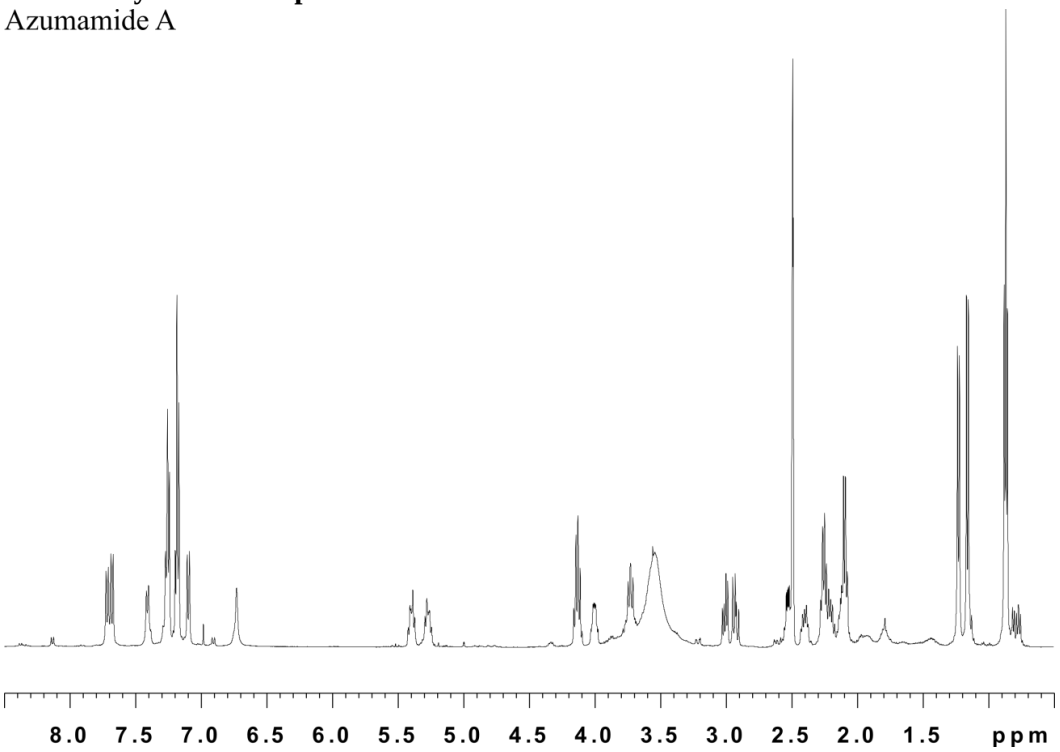



Epi-10,23-dihydro-24,25-dehydroaflavine

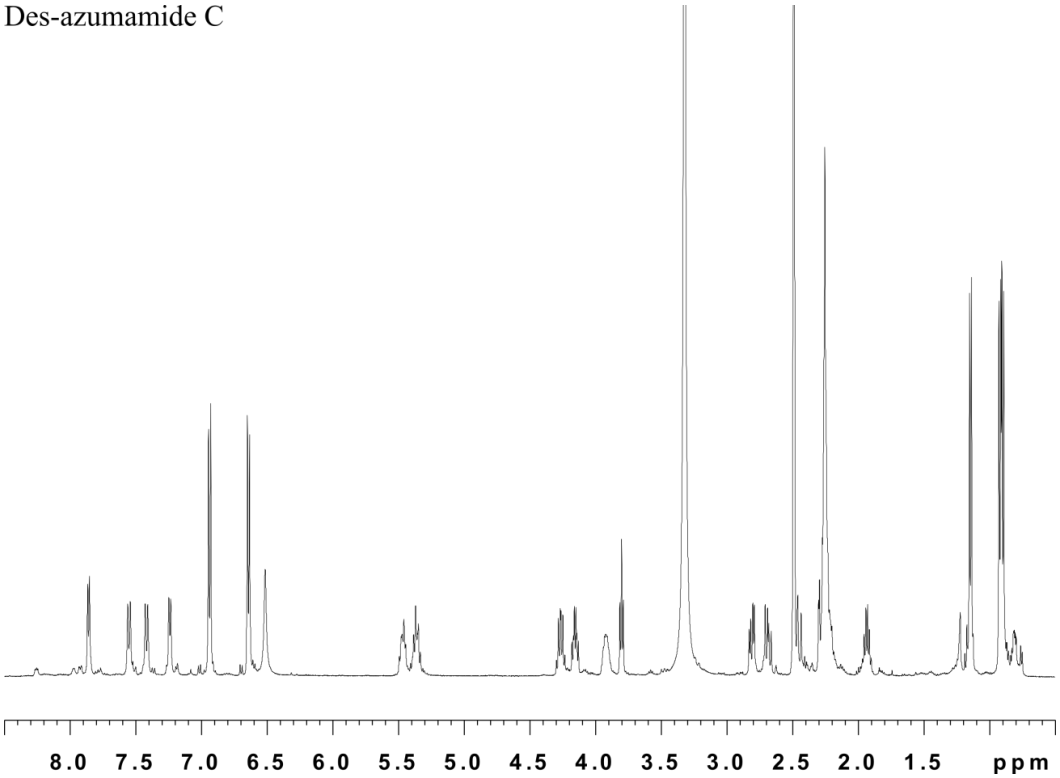


### A5.1.2. Synthetic compounds

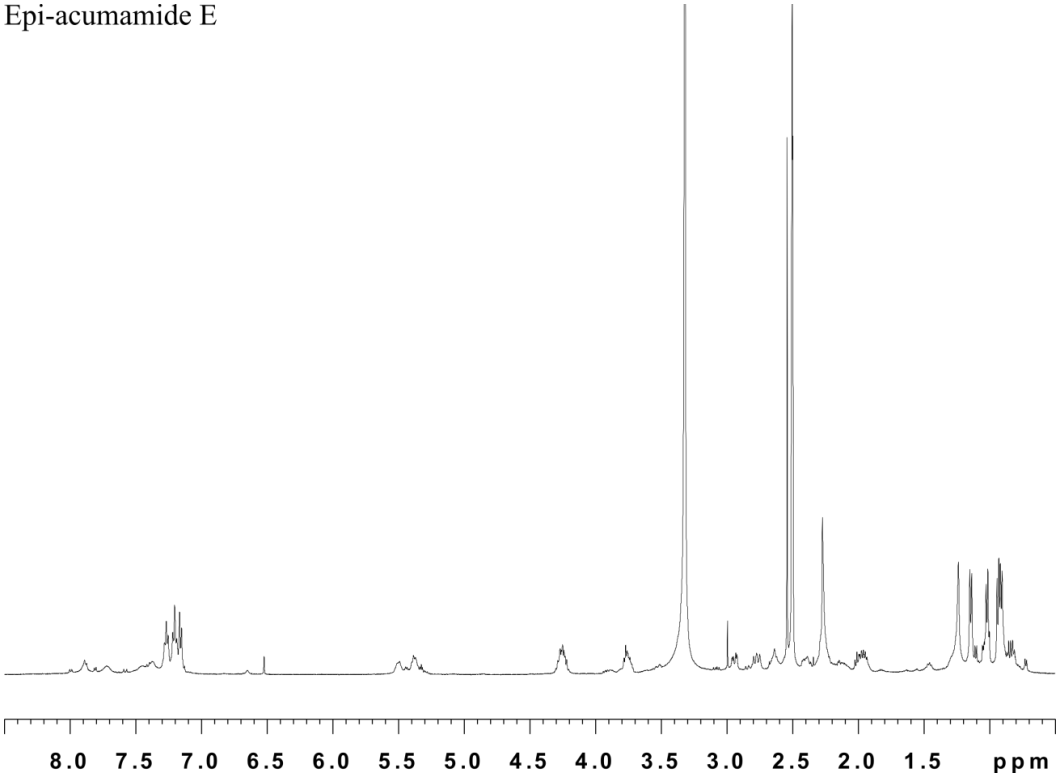
Azumamide A



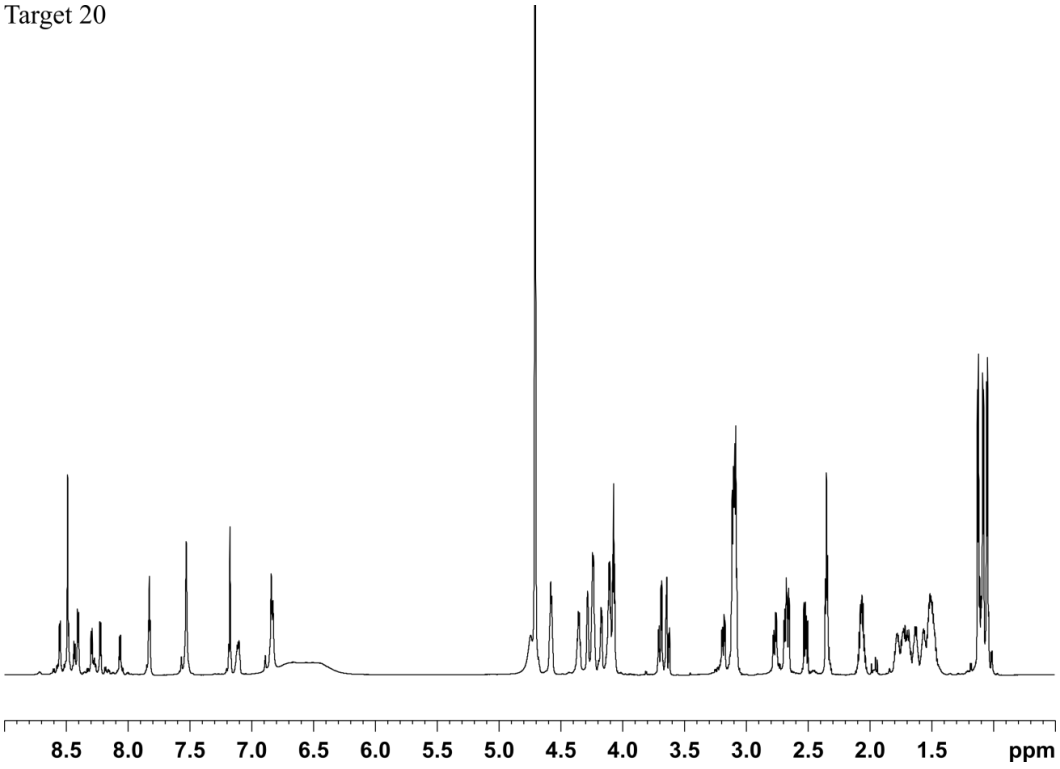

Des-azumamide C



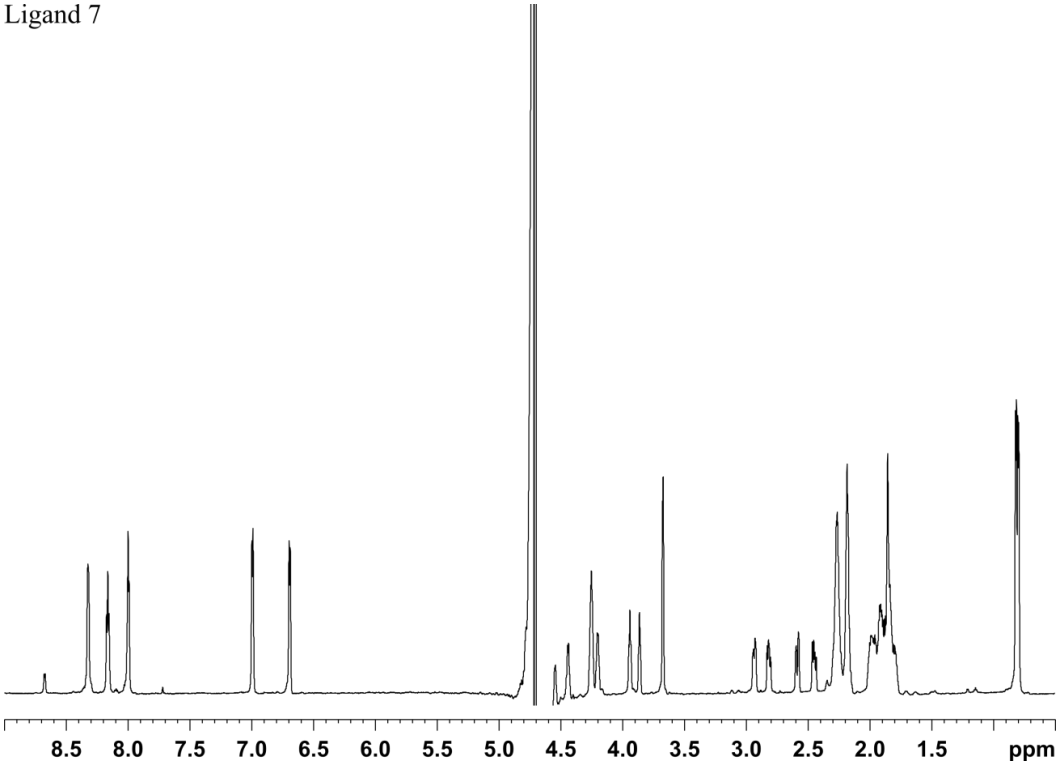
Epi-acumamide E



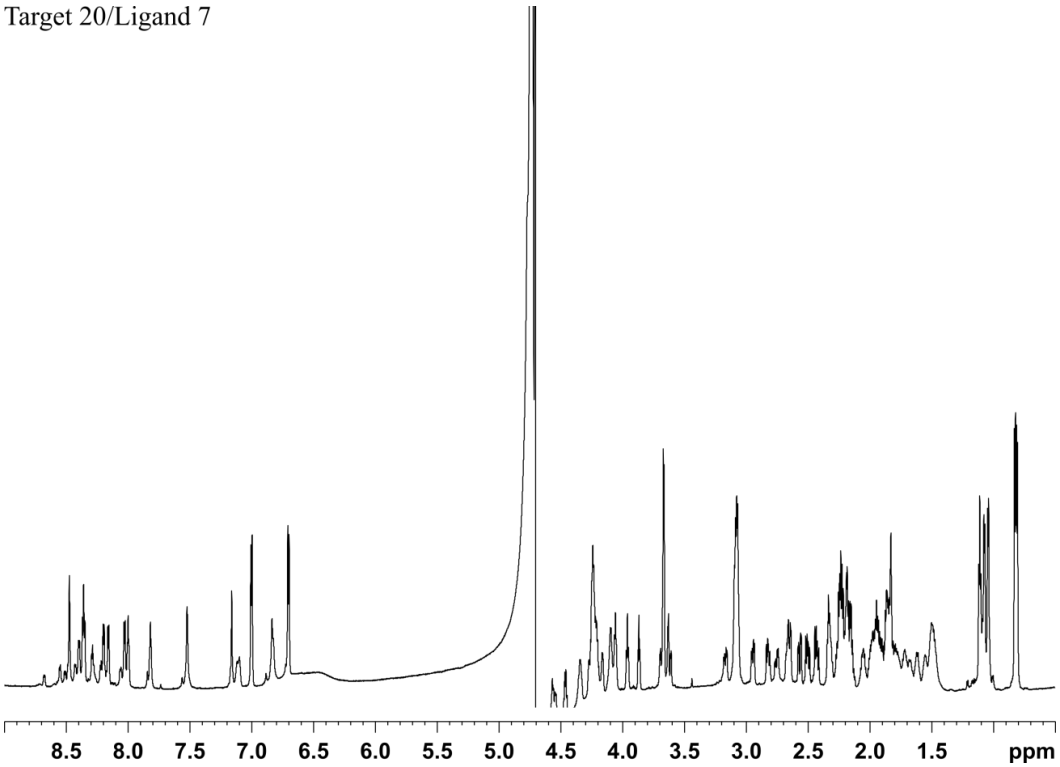

Target 20



Ligand 7

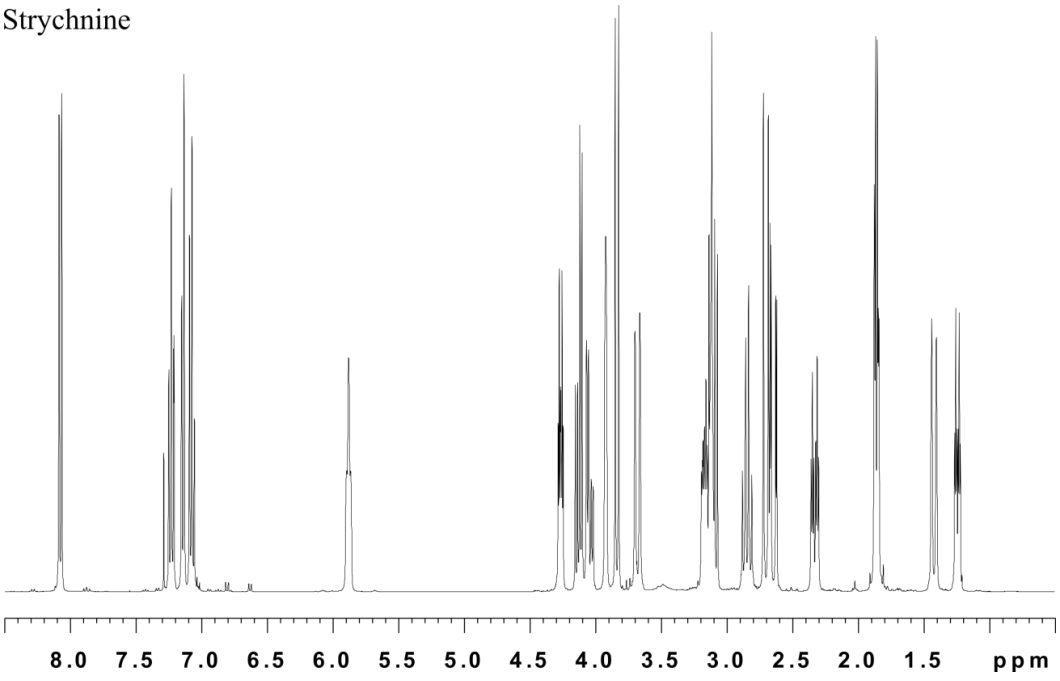



Target 20/Ligand 7

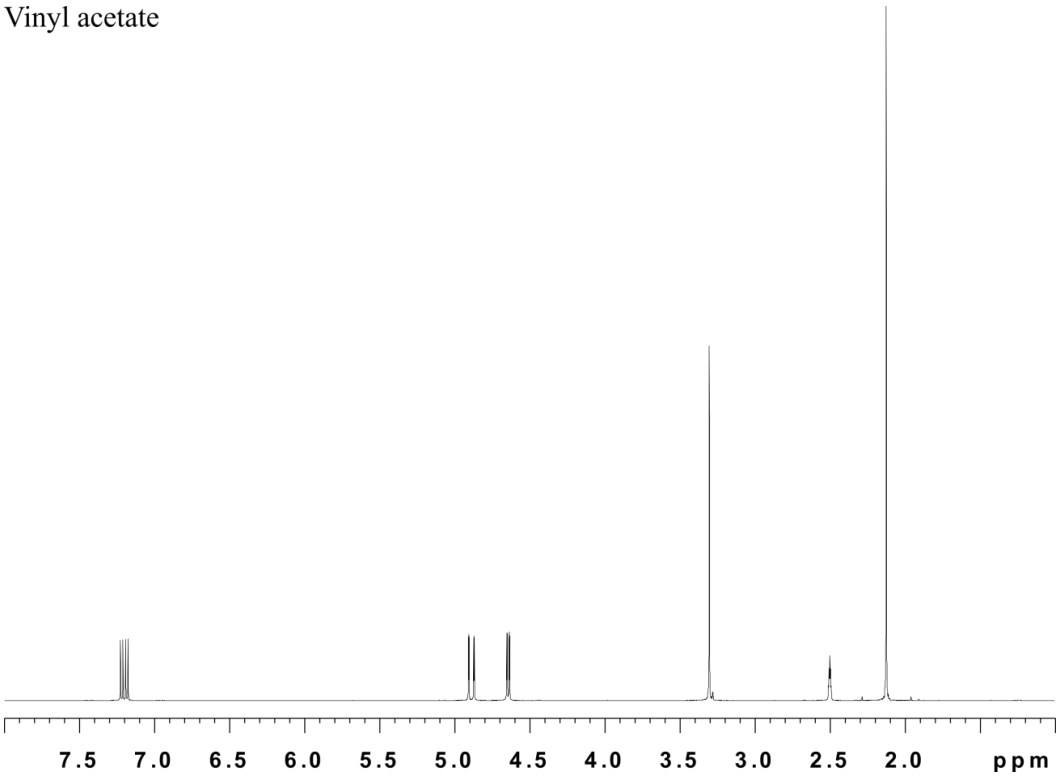


### A5.1.3. $S^3$ HMBC

Strychnine

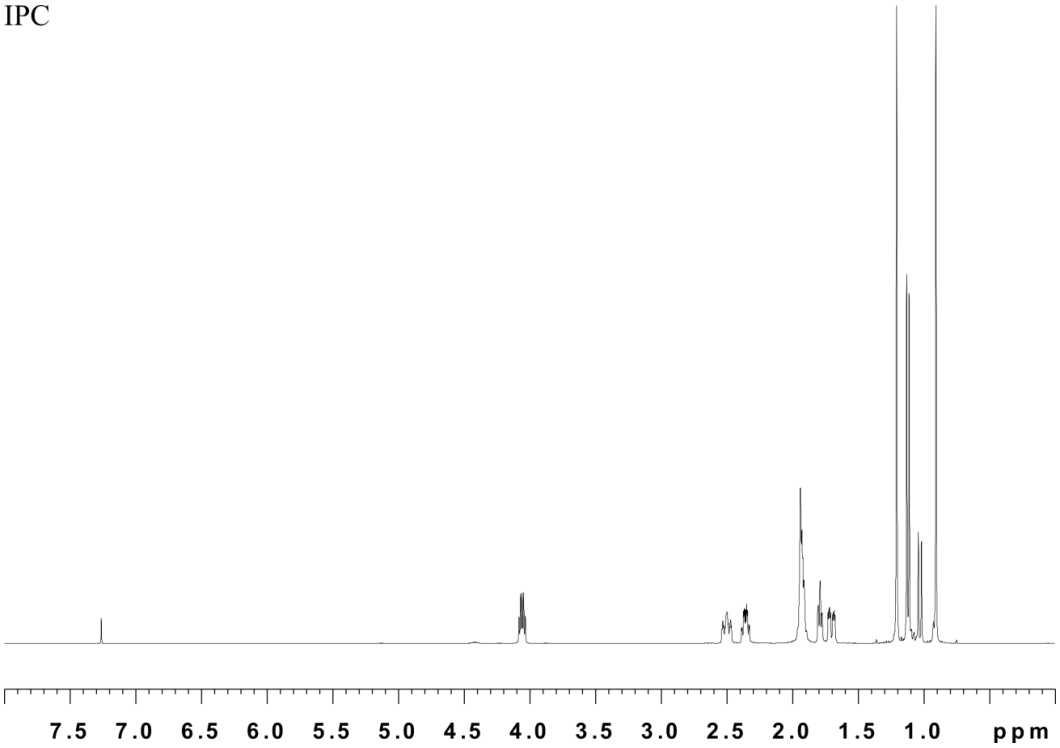



Vinyl acetate

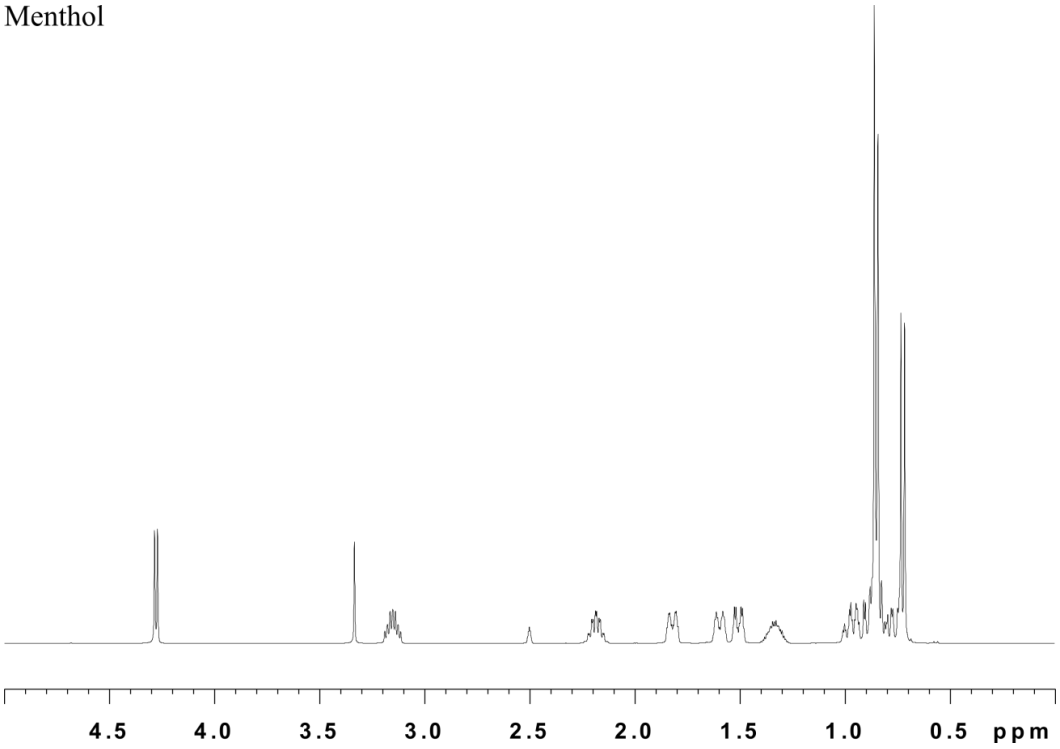


### A5.1.4. RDC

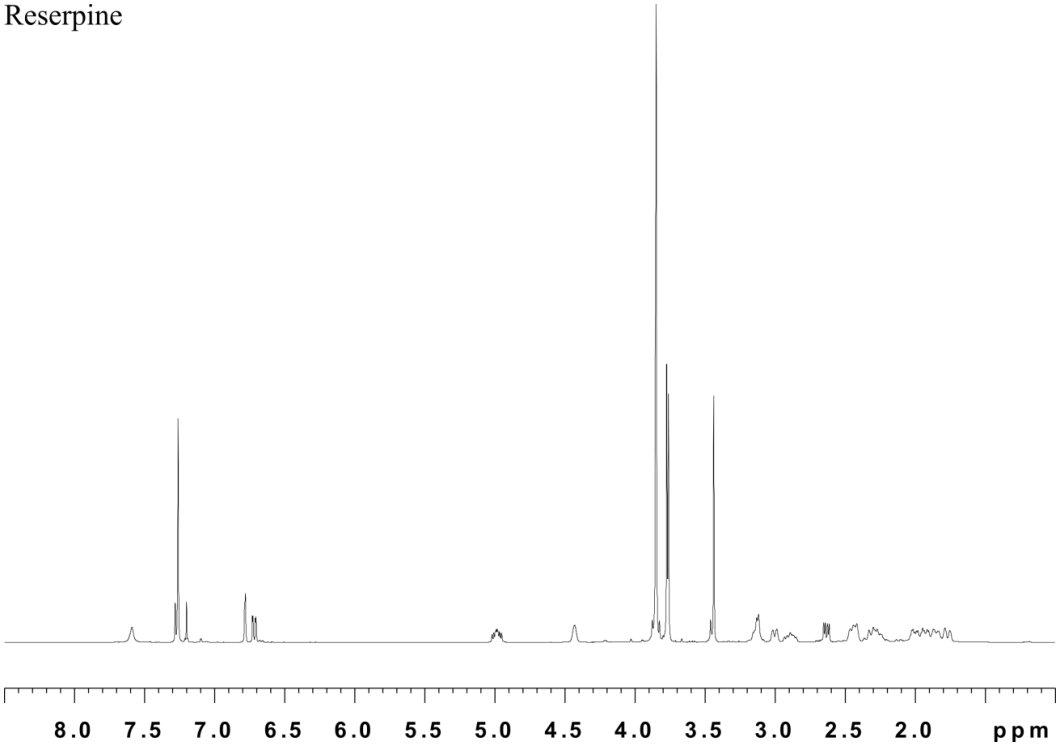
IPC



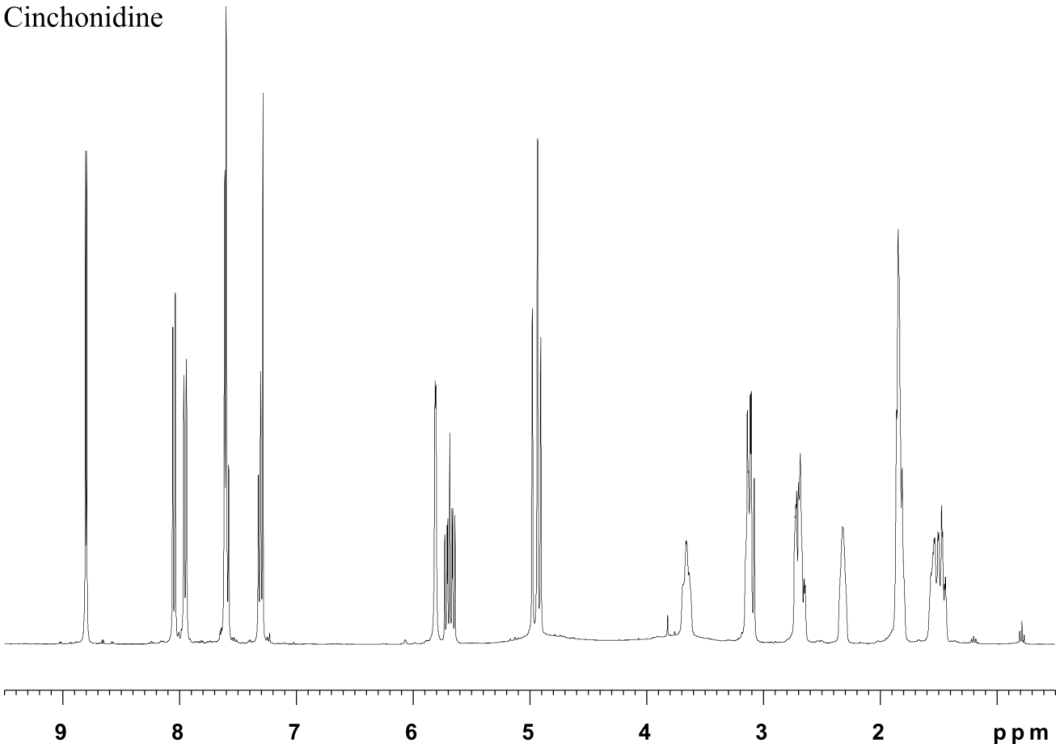

Menthol



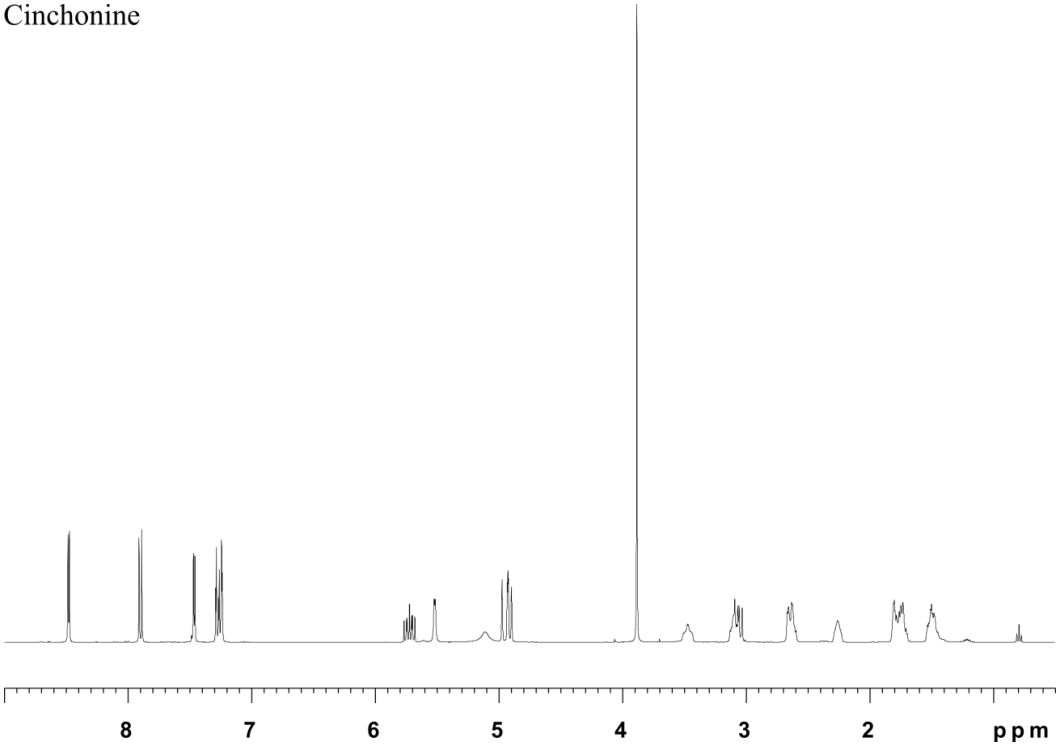
Reserpine



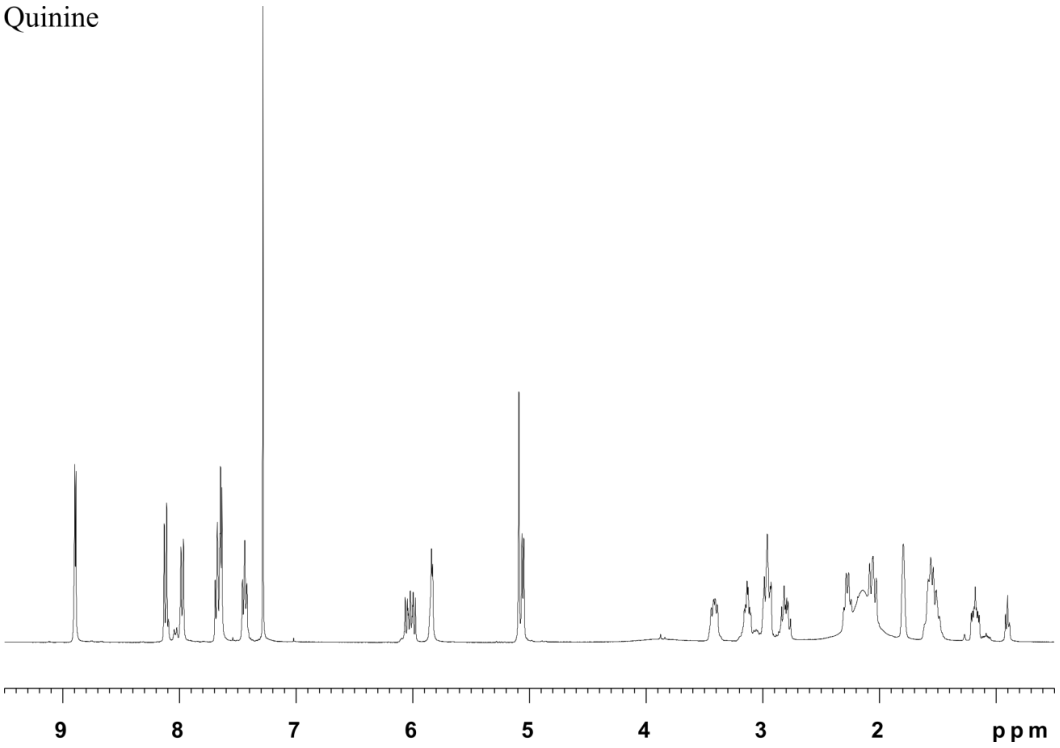

Cinchonidine



Cinchonine

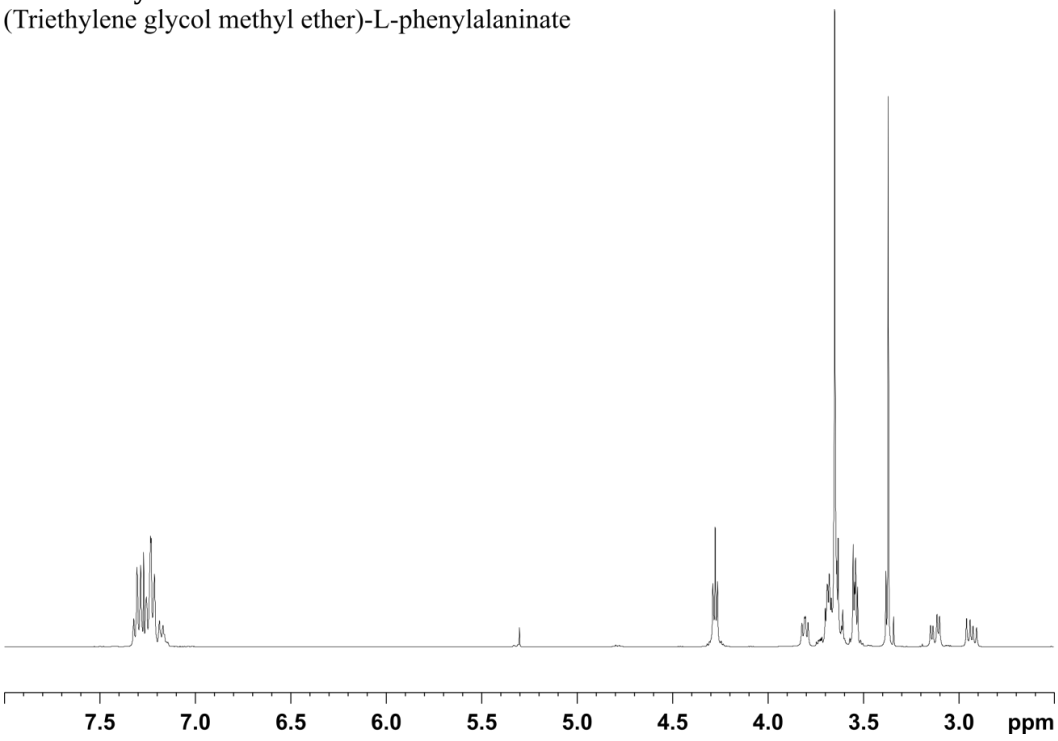



Quinine

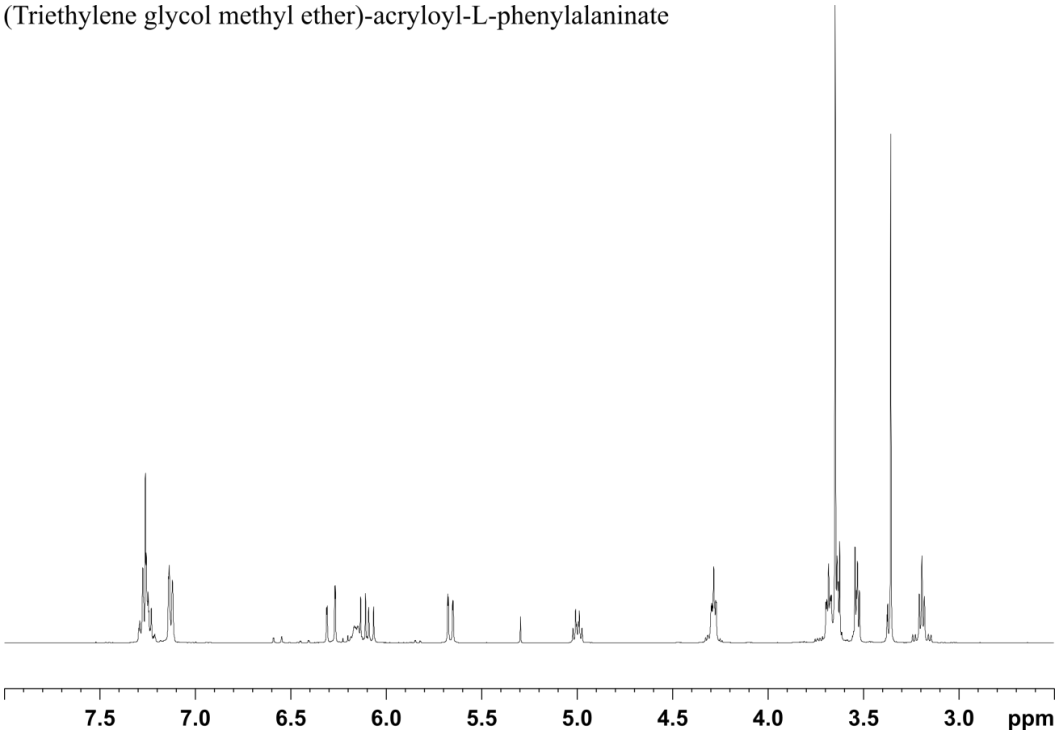


### A5.1.5. Synthesized monomers

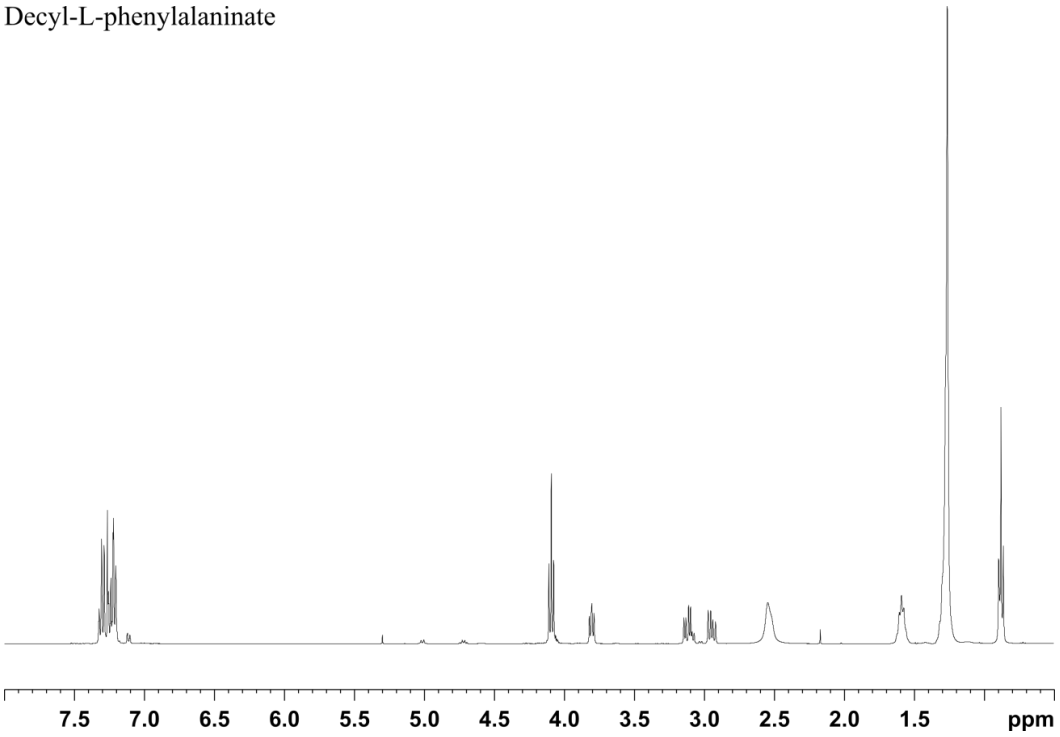
(Triethylene glycol methyl ether)-L-phenylalaninate



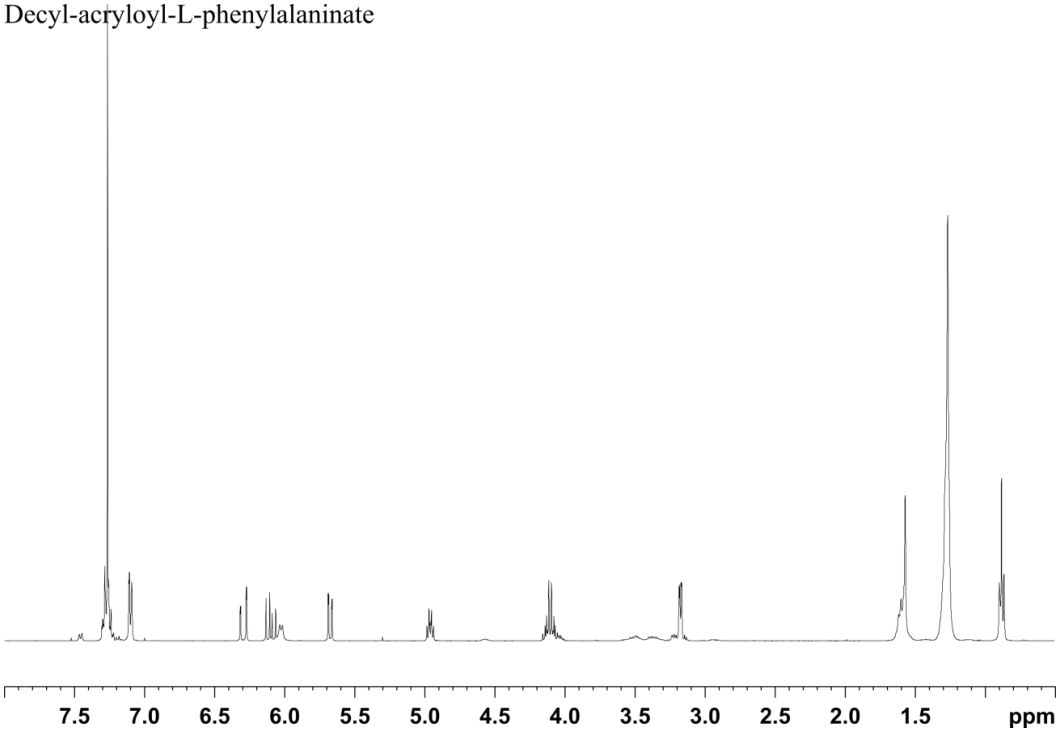

(Triethylene glycol methyl ether)-acryloyl-L-phenylalaninate



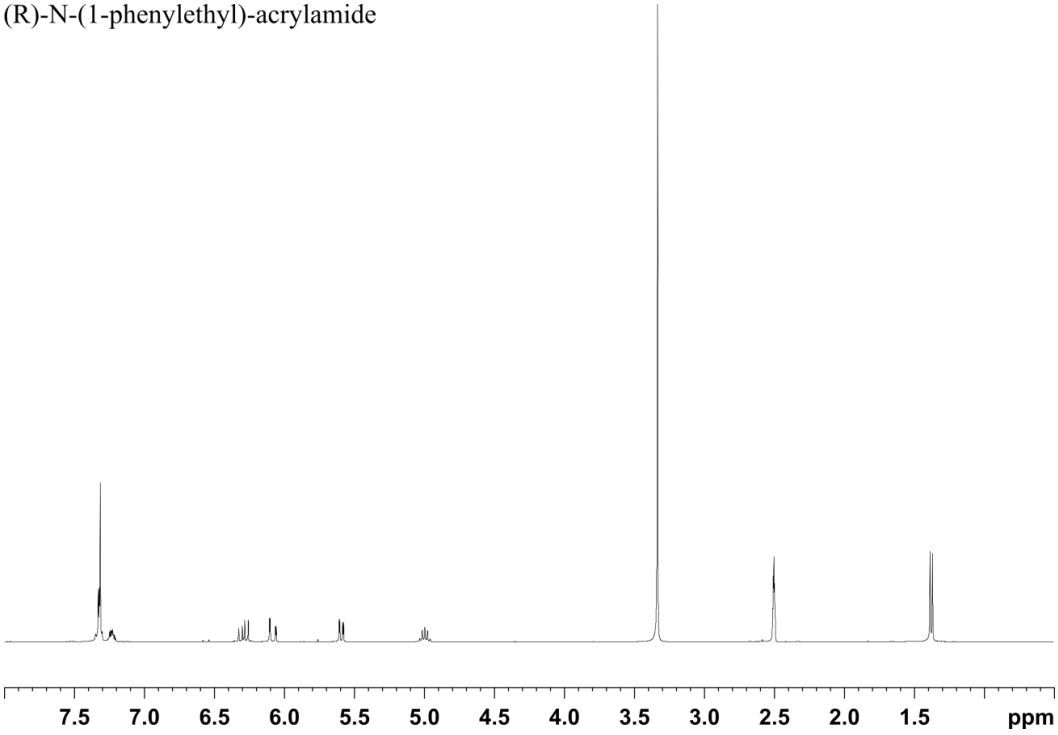
Decyl-L-phenylalaninate



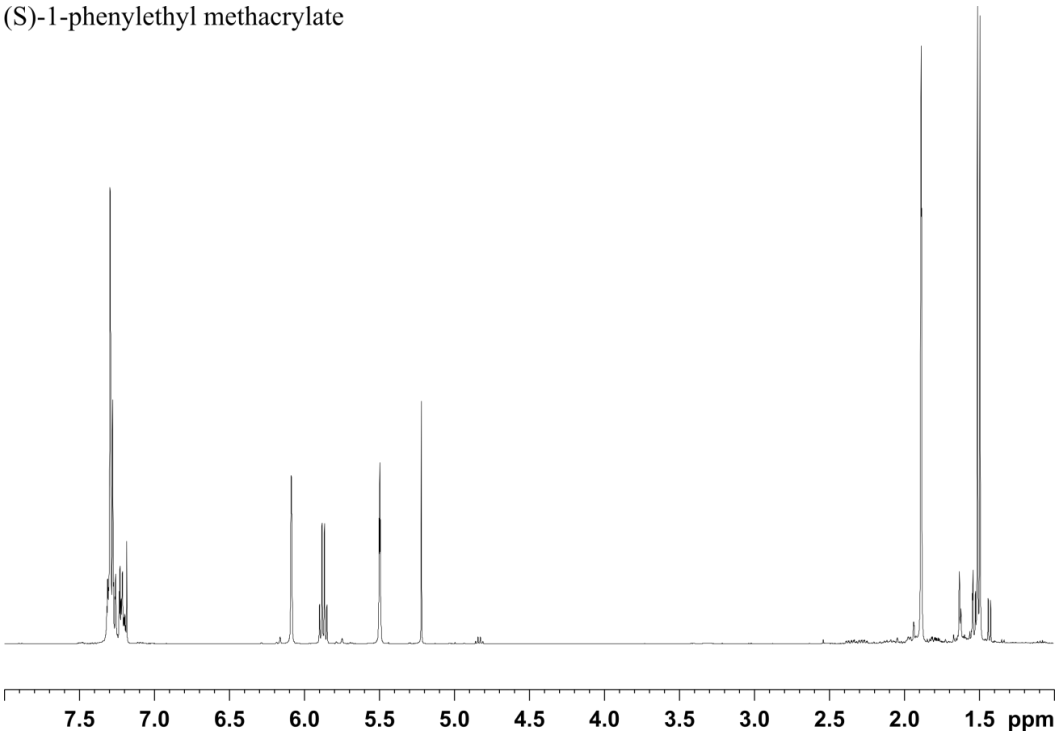

Decyl-acryloyl-L-phenylalaninate



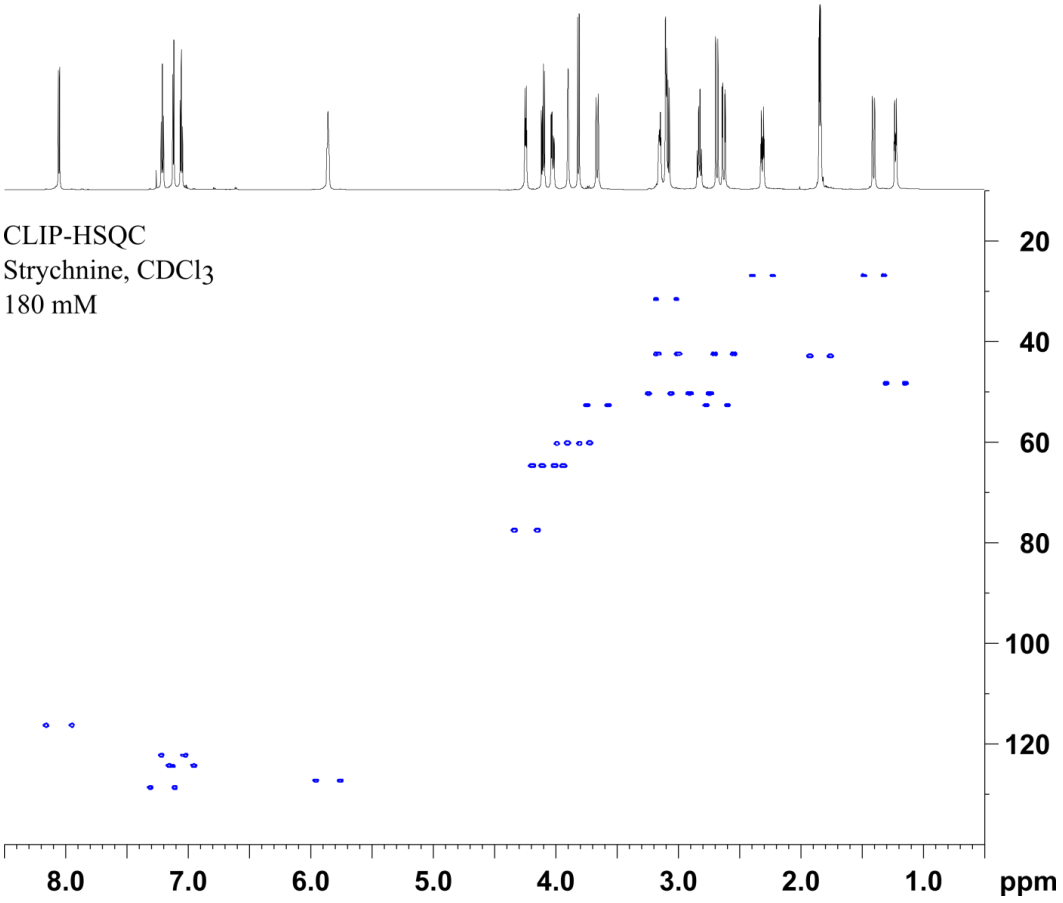
(R)-N-(1-phenylethyl)-acrylamide

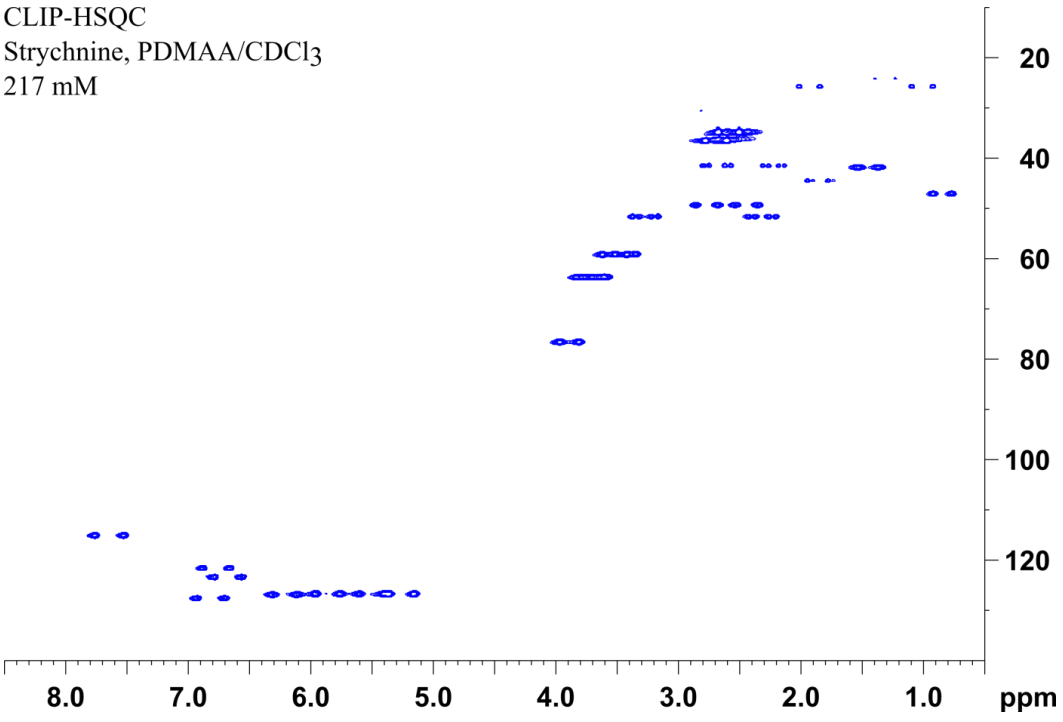


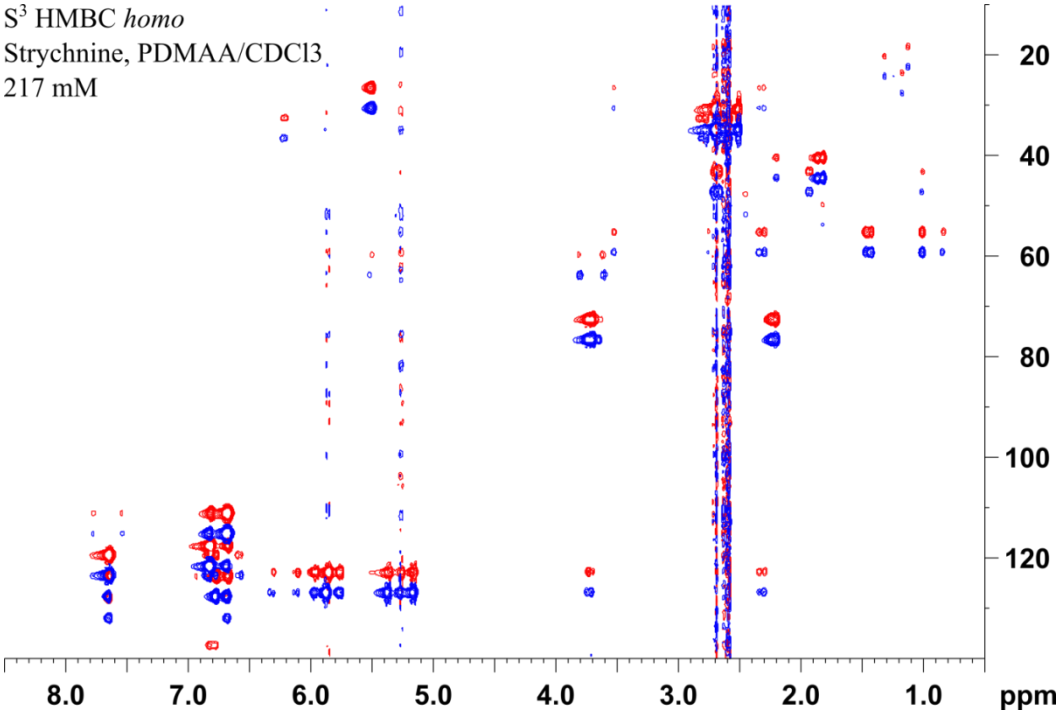

(S)-1-phenylethyl methacrylate



## A5.2. Examples of isotropic and aligned spectra




S3 HMBC homo
Strychnine, PDMAA/CDCl3
217 mM
20
40
60
80
100
120
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
ppm

# A6. $S^3$ HMBC

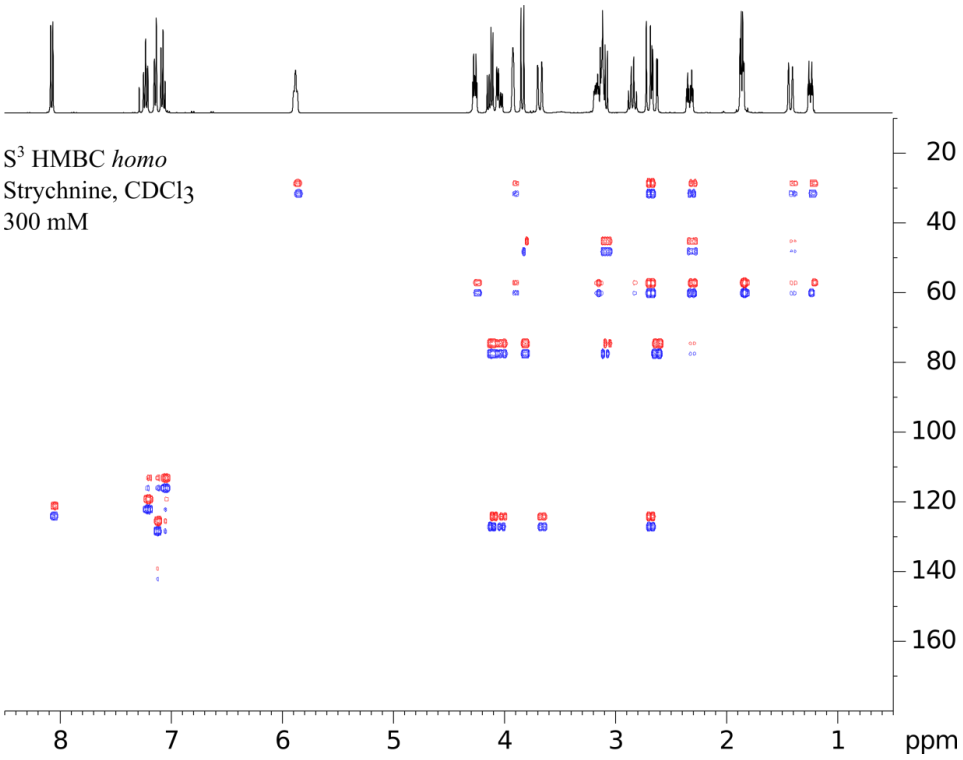



**Figure A6.1. $CH/CH_3$ $S^3$ HMBC homo spectra of strychnine. Add spectrum in blue and subtract spectrum in red. Subtract spectrum is offset in F1 dimension.**

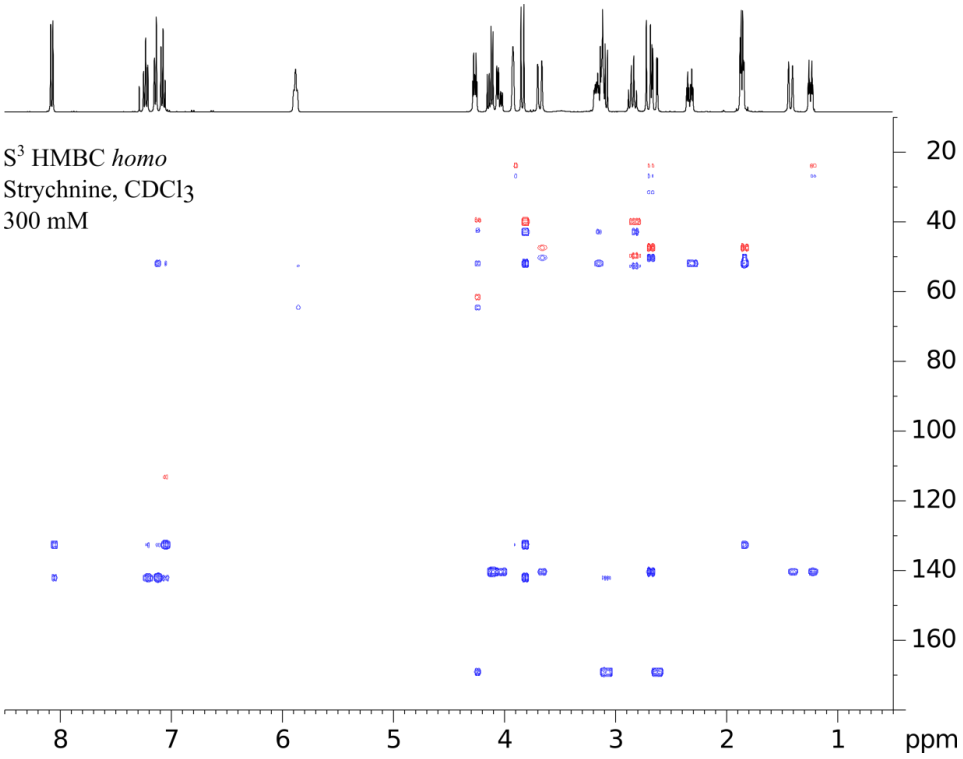



**Figure A6.2. $CH_2/C_q$ $S^3$ HMBC homo spectra of strychnine. Add spectrum in blue and subtract spectrum in red. Subtract spectrum is offset in F1 dimension.**

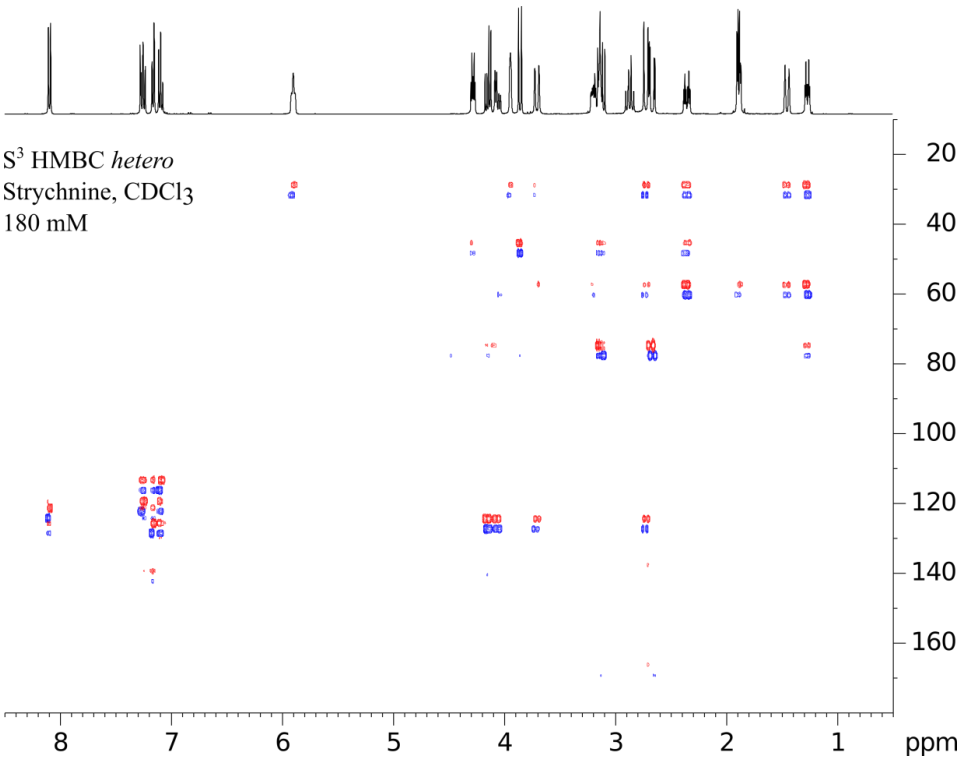



**Figure A6.3. $CH/CH_3$ $S^3$ HMBC hetero spectra of strychnine. Add spectrum in blue and subtract spectrum in red. Subtract spectrum is offset in F1 dimension.**

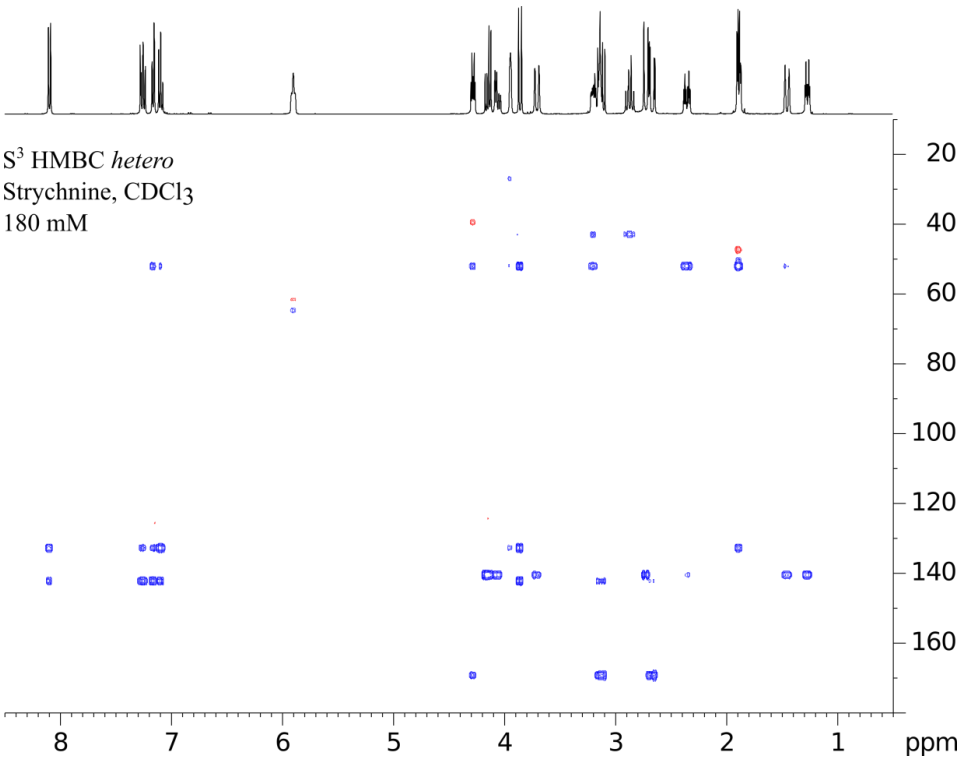



**Figure A6.4. $CH_2/C_q$ $S^3$ HMBC hetero spectra of strychnine. Add spectrum in blue and subtract spectrum in red. Subtract spectrum is offset in F1 dimension.**

| Nuc 1 | Nuc 2 | Exp | SVD | Exp | SVD | SVD scaled |
|---|---|---|---|---|---|---|
| C1 | H1 | 19.0 | 18.0 | 19.0 | 16.5 | 17.6 |
| C2 | H2 | 14.0 | 14.8 | 14.0 | 15.3 | 14.3 |
| C3 | H3 | 18.8 | 18.2 | 18.8 | 16.9 | 17.2 |
| C4 | H4 | 18.7 | 18.3 | 18.7 | 16.8 | 17.9 |
| C8 | H8 | -11.0 | -10.3 | -11.0 | -8.0 | -9.9 |
| C11 | H11a | 5.1 | 7.5 | 5.1 | 5.4 | 7.5 |
| C11 | H11b | -21.7 | -23.7 | -21.7 | -21.3 | -22.8 |
| C12 | H12 | -25.2 | -25.5 | -25.2 | -25.6 | -24.6 |
| C13 | H13 | -3.8 | -3.6 | -3.8 | -3.0 | -3.5 |
| C14 | H14 | -25.1 | -24.2 | -25.1 | -24.3 | -23.4 |
| C15 | H15a | 3.1 | 3.2 | 3.1 | 3.6 | 3.1 |
| C15 | H15b | 6.1 | 5.9 | 6.1 | 5.4 | 5.6 |
| C16 | H16 | 15.0 | 14.1 | 15.0 | 14.7 | 13.8 |
| C18 | H18a | -2.9 | -4.3 | -2.9 | -6.1 | -4.0 |
| C18 | H18b | 13.2 | 13.6 | 13.2 | 14.3 | 13.3 |
| C20 | H20a | -14.4 | -12.5 | -14.4 | -10.4 | -12.0 |
| C20 | H20b | -3.5 | -3.1 | -3.5 | -2.8 | -3.1 |
| C22 | H22 | 0.1 | 2.2 | 0.1 | 1.7 | 2.0 |
| H1 | H2 | - | 6.1 | 3.9 | 5.6 | 5.7 |
| H1 | H3 | - | 1.2 | 1.6 | 1.1 | 1.2 |
| H2 | H3 | - | 6.0 | 3.4 | 5.6 | 5.9 |
| H2 | H4 | - | 1.0 | 0.3 | 1.0 | 1.0 |
| H3 | H4 | - | 4.9 | 4.6 | 5.0 | 4.7 |
| H8 | H12 | - | -2.0 | -3.3 | -2.0 | -1.9 |
| H8 | H13 | - | -5.5 | -4.2 | -5.2 | -5.3 |
| H8 | H16 | - | 0.6 | 2.1 | 0.7 | 0.6 |
| H11b | H12 | - | -5.7 | -3.2 | -5.3 | -5.5 |
| H12 | H23a | - | 1.0 | 1.5 | 1.0 | 1.0 |
| H12 | H23b | - | 7.8 | 9.4 | 7.1 | 7.4 |
| H13 | H14 | - | 4.3 | 2.8 | 3.9 | 4.0 |
| H13 | H15a | - | 1.9 | 2.1 | 1.8 | 1.9 |
| H13 | H15b | - | 5.9 | 8.1 | 6.0 | 5.8 |
| H14 | H15b | - | 4.5 | 3.8 | 4.5 | 4.3 |
| H14 | H16 | - | 0.2 | 1.1 | 0.1 | 0.2 |
| H14 | H20b | - | -0.6 | -0.4 | -0.7 | -0.6 |
| H14 | H22 | - | -0.9 | -0.7 | -0.7 | -0.8 |
| H15a | H16 | - | -4.4 | -3.0 | -4.5 | -4.3 |
| H15b | H16 | - | -3.7 | -2.5 | -4.2 | -3.5 |
| H16 | H17a/H17b | - | 4.8 | -4.1 | -4.54* | -4.76* |
| H16 | H18a | - | -1.8 | -1.7 | -1.6 | -1.7 |
| H16 | H20b | - | 1.0 | 1.9 | 1.1 | 1.0 |
| H20b | H22 | - | 5.6 | 5.4 | 5.4 | 5.5 |
| H22 | H23a | - | 5.2 | 5.3 | 4.5 | 5.1 |
| H22 | H23b | - | -1.2 | -2.1 | -1.5 | -1.2 |
| | | | *Q*-factor | 0.080 | 0.149 | 0.126 |
| | | | $D_a$ | 7.293e-04 | 6.937e-04 | 7.021e-04 |

**Table A6-1. Experimental data from CLIP-HSQC and $S^3$ HMBC *homo* of strychnine compared to back-calculated RDCs using SVD or SVD Hz scaling in the program MSPIN. When no experimental data (-) is given the theoretical values are calculated from the existing data. *The values were treated as the average and not included in the back-calculated fit, though the values are included in the *Q*-factor. Polymer 6.2.**

## A6.1. Automatic bias-free S$^3$ HMBC *J*-coupling extraction

The script package rbnmr (Nils Nyberg, Copenhagen University, 2013) was used to import data from Topspin files into Matlab. Two scripts were written in Matlab to aid in the extraction of couplings.

The first will determine a coupling constant, given an input row number and peak boundaries. The second script will find peaks and determine couplings from the 2D spectra and consists of two main parts; a module to locate and match the involved peaks and a module which extract the coupling constants.

The script will first read the 2D data spectra and identify the peaks involved in couplings. The row number with the coupling constants(s), defined as the highest intensity for the resonances of a given constant, will be outputted to help in deciding which rows to manually extract if intended. Data will be saved as "1D" data slices defining the individual peaks involved in couplings. The second module will then match the peaks from the two spectra and overlay these. The overlay is scored by equation (6.1), the minimum is identified and the displacement given in Hz.

$$\sum_i |p_{+,i} - p_{-,i}| + |p_{+,max} - p_{-,j}| + |p_{+,k} - p_{-,max}| \qquad \textbf{(6.1)}$$

Where *p* are points in the 1D slices from the + or – spectra, *i* is the chemical shift of *p*, *j*,*k* are the shift values of the maximum intensity of the partnering (+ for – and *vice versa*) slice. The max terms of the score were included due to necessity; if a peak is slightly br.er than its partner, the fit may be equally good for multiple positions. This is avoided by scoring the points of maximum intensity higher. Harsher scoring of the maximum points assumes that the fit is better for higher intensities, which may not be true. The approach was found to lead to a more intuitive fit where the peaks are always centered, and in by far the most cases the fit is completely independent of the term.

The peaks are displaced by keeping $p_{+,i}$ constant in equation (6.1) and setting $p_{-,i} = p_{+,i\pm1}$. The spacing in between $p_i$ and $p_{i+1}$ is equal to the theoretical minimum detectable coupling constant for the script and may be lowered by linear interpolation of the data, given the data has a reasonable F2 resolution. The value is usually set to a minimum of 0.01 Hz which should be well below experimental errors. Typically a cutoff of 95 % of the maximum intensities was set for the peaks, so that the only the upper 95 % was used in the fitting.

Figures of the overlays (not used for figures in the article) are made to investigate the result, where mistakes are immediately obvious as only the parts of the spectra used in the fitting is plotted. Lastly the couplings may be matched by their H- and C-shifts to a given atom numbering, if provided, to ease further analysis. Care should be taken when overlap of peaks are present, just as with manual extraction. Also coupling patterns with baseline separation in between peaks may be taken for two or more couplings. Here the second script may be used to manually give the boundaries and thus avoid or minimize problems.

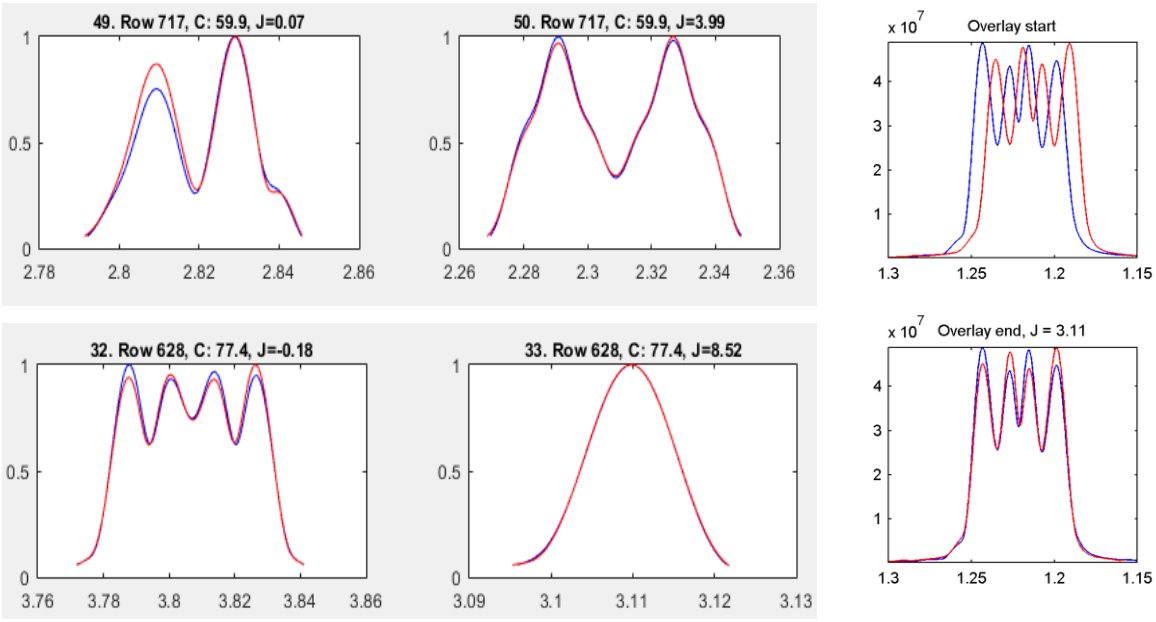



**Figure A6.5. Example of output of the Matlab® script. Left: fully automated identification and extraction of coupling constants. Accessed by overlay. Rigt: Manual regions given as input (row, minimum- and maximum $\delta_H$). The minimum differentiation is 0.5 Hz, compared to 0.1 Hz used throughout the thesis. Exemplified by $S^3$ HMBC homo cross-peaks.**

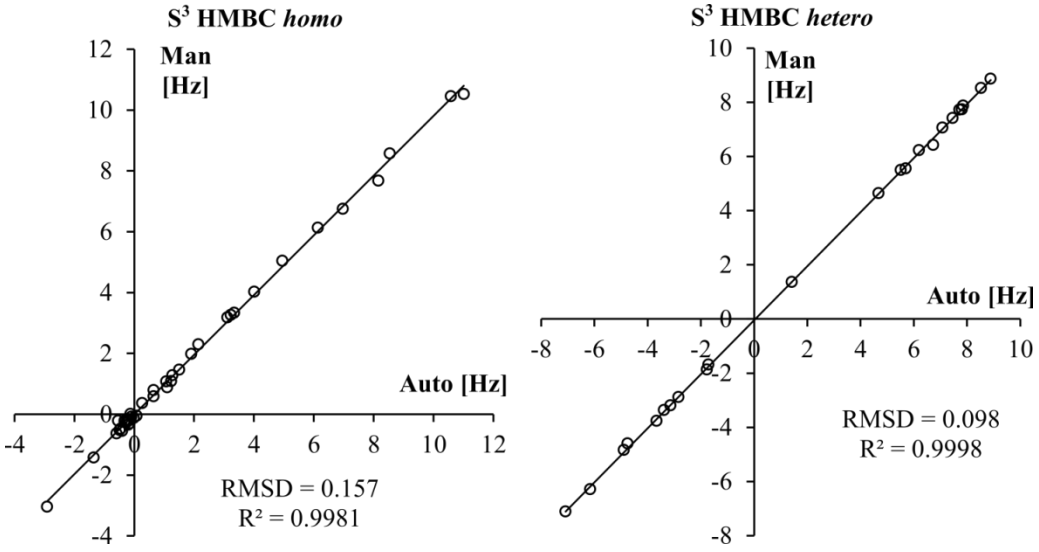



**Figure A6.6. Correlation of coupling constants obtained from the Matlab® script and those manually extracted for the two $S^3$ HMBCs.**

## A6.2. Heteronuclear coupling constant DFT

**Table A6-2. Comparison of experimental and theoretical J-coupling constants calculated as given. For each dataset the unscaled is given in the left column and a scaled set in the right. Scaled by linear fit to experimental data (a and b given)**

| | | | Opt: | b3lyp/6-31g(d) | | b3lyp/6-31g(d) | | b3lyp/6-31g(d) | | b3lyp/6-31g(d) | | b3lyp/6-31g(d) | | b3lyp/6-31g(d) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C | H | Exp | J-calc | b3lyp/6-31g(d) | | b3lyp/6-311g(d,p) | | b3lyp/6-311+g(d,p) | | mpw1pw91/6-31g(d) | | mpw1pw91/6-311g(d,p) | | mpw1pw91/6-311+g(d,p) | |
| 1 | 3 | 7.4 | | 6 (1.4) | 7.2 (0.2) | 6.6 (0.8) | 7.4 (0) | 6.6 (0.8) | 7.4 (0.1) | 6.4 (1) | 7.4 (0) | 6.7 (0.7) | 7.5 (0) | 6.7 (0.7) | 7.5 (0) |
| 2 | 4 | 7.9 | | 6.2 (1.7) | 7.4 (0.4) | 6.9 (0.9) | 7.7 (0.2) | 6.9 (1) | 7.7 (0.2) | 6.6 (1.2) | 7.7 (0.2) | 7.1 (0.8) | 7.8 (0.1) | 7 (0.8) | 7.8 (0.1) |
| 3 | 1 | 8.9 | | 7.2 (1.7) | 8.9 (0) | 8 (0.9) | 8.9 (0.1) | 8 (0.9) | 9 (0.1) | 7.6 (1.3) | 8.9 (0.1) | 8.1 (0.8) | 8.9 (0) | 8.1 (0.8) | 8.9 (0.1) |
| 4 | 2 | 7.7 | | 6.2 (1.5) | 7.5 (0.2) | 7 (0.7) | 7.8 (0.1) | 7 (0.7) | 7.8 (0.1) | 6.6 (1.1) | 7.7 (0) | 7.1 (0.6) | 7.9 (0.1) | 7.1 (0.7) | 7.8 (0.1) |
| 8 | 12 | 6.4 | | 5.6 (0.9) | 6.5 (0.1) | 5.7 (0.7) | 6.3 (0.1) | 5.6 (0.8) | 6.2 (0.2) | 5.5 (0.9) | 6.2 (0.2) | 5.6 (0.8) | 6.2 (0.2) | 5.5 (0.9) | 6.2 (0.3) |
| 8 | 13 | -1.7 | | 1.3 (2.9) | 0.1 (1.8) | -0.7 (1) | -1.1 (0.6) | -0.9 (0.8) | -1.3 (0.3) | 0.5 (2.2) | -0.5 (1.2) | -1.4 (0.3) | -1.4 (0.3) | -1.6 (0.1) | -1.5 (0.2) |
| 11b | 12 | -7.1 | | -4.3 (2.9) | -8.2 (1.1) | -6.3 (0.8) | -7.6 (0.5) | -6.3 (0.8) | -7.6 (0.5) | -4.8 (2.3) | -7.5 (0.4) | -6.9 (0.2) | -7.4 (0.2) | -6.9 (0.2) | -7.3 (0.2) |
| 13 | 8 | -6.3 | | -2.4 (3.9) | -5.4 (0.9) | -5.3 (1) | -6.4 (0.1) | -5.4 (0.9) | -6.5 (0.2) | -3.1 (3.2) | -5.2 (1.1) | -6 (0.3) | -6.3 (0) | -6 (0.3) | -6.4 (0.1) |
| 13 | 14 | -4.9 | | -2.5 (2.3) | -5.6 (0.8) | -4.2 (0.7) | -5.1 (0.3) | -4.2 (0.7) | -5.1 (0.3) | -3.3 (1.5) | -5.5 (0.7) | -4.8 (0.1) | -5 (0.2) | -4.8 (0.1) | -5 (0.2) |
| 15a | 13 | 7.7 | | 6.8 (1) | 8.3 (0.6) | 7.1 (0.6) | 7.9 (0.2) | 7.1 (0.6) | 8 (0.2) | 6.8 (0.9) | 7.9 (0.2) | 7 (0.8) | 7.7 (0) | 7 (0.7) | 7.7 (0) |
| 15a | 14 | -2.9 | | 0.1 (3) | -1.7 (1.2) | -1.9 (1) | -2.5 (0.4) | -1.9 (1) | -2.5 (0.4) | -0.5 (2.4) | -1.8 (1.1) | -2.5 (0.4) | -2.6 (0.3) | -2.5 (0.4) | -2.6 (0.3) |
| 15a | 16 | -4.6 | | -1.6 (3) | -4.2 (0.4) | -3.7 (0.9) | -4.6 (0) | -3.7 (1) | -4.5 (0.1) | -2.3 (2.4) | -4.1 (0.5) | -4.4 (0.2) | -4.6 (0) | -4.3 (0.3) | -4.5 (0.1) |
| 15b | 14 | -3.2 | | -1 (2.2) | -3.4 (0.2) | -2.5 (0.7) | -3.1 (0.1) | -2.4 (0.8) | -3.1 (0.1) | -1.7 (1.5) | -3.4 (0.2) | -3.1 (0.1) | -3.2 (0) | -3.1 (0.1) | -3.2 (0) |
| 15b | 16 | -1.9 | | 0.4 (2.3) | -1.2 (0.7) | -1.4 (0.5) | -1.9 (0) | -1.3 (0.5) | -1.8 (0) | -0.4 (1.4) | -1.7 (0.2) | -2.2 (0.3) | -2.2 (0.3) | -2.1 (0.3) | -2.2 (0.3) |
| 16 | 14 | 6.2 | | 5.5 (0.7) | 6.4 (0.2) | 5.9 (0.3) | 6.5 (0.3) | 6 (0.3) | 6.6 (0.4) | 5.5 (0.8) | 6.1 (0.1) | 5.7 (0.5) | 6.4 (0.1) | 5.8 (0.5) | 6.4 (0.2) |
| 20a | 14 | 1.4 | | 0.9 (0.5) | -0.5 (1.8) | 1.1 (0.3) | 0.9 (0.4) | 1.1 (0.3) | 0.9 (0.4) | 1.1 (0.3) | 0.3 (1.1) | 1.3 (0.1) | 1.5 (0.2) | 1.3 (0.1) | 1.5 (0.2) |
| 20a | 22 | 5.6 | | 5.9 (0.4) | 7 (1.5) | 4.4 (1.2) | 4.8 (0.8) | 4.4 (1.2) | 4.8 (0.8) | 6.7 (1.2) | 7.8 (2.2) | 4.4 (1.1) | 5 (0.6) | 4.4 (1.1) | 5 (0.6) |
| 20b | 14 | 5.5 | | 4.1 (1.4) | 4.3 (1.2) | 4.7 (0.8) | 5.1 (0.4) | 4.7 (0.8) | 5.1 (0.4) | 4.2 (1.3) | 4.4 (1.1) | 4.7 (0.8) | 5.2 (0.3) | 4.6 (0.9) | 5.2 (0.3) |
| 20b | 16 | 7.1 | | 5.9 (1.2) | 6.9 (0.1) | 6.3 (0.7) | 7 (0) | 6.3 (0.8) | 7 (0.1) | 5.9 (1.2) | 6.6 (0.4) | 6.2 (0.9) | 6.9 (0.2) | 6.1 (0.9) | 6.8 (0.3) |
| 20b | 22 | 4.6 | | 4.5 (0.2) | 4.9 (0.2) | 4.7 (0.1) | 5.2 (0.5) | 4.8 (0.1) | 5.2 (0.6) | 4.4 (0.2) | 4.7 (0.1) | 4.5 (0.1) | 5.1 (0.4) | 4.6 (0.1) | 5.1 (0.5) |
| 22 | 14 | 8.5 | | 6.9 (1.7) | 8.4 (0.1) | 7.7 (0.8) | 8.7 (0.1) | 7.7 (0.8) | 8.7 (0.1) | 7.3 (1.2) | 8.5 (0) | 8 (0.5) | 8.8 (0.3) | 8 (0.5) | 8.8 (0.3) |
| 23a | 22 | -3.8 | | -2.2 (1.6) | -5 (1.3) | -3.1 (0.7) | -3.9 (0.1) | -3.1 (0.7) | -3.9 (0.1) | -3 (0.8) | -5 (1.3) | -3.8 (0) | -4 (0.2) | -3.8 (0) | -3.9 (0.2) |
| 23b | 22 | -3.4 | | -1.2 (2.2) | -3.6 (0.2) | -2 (1.4) | -2.6 (0.7) | -2 (1.3) | -2.7 (0.7) | -2.4 (1) | -4.2 (0.9) | -2.9 (0.5) | -3 (0.4) | -2.9 (0.5) | -3 (0.4) |
| | | | a | 0.7 | | 0.9 | | 0.9 | | 0.8 | | 0.9 | | 0.9 | |
| | | | b | 1.2 | | 0.3 | | 0.3 | | 0.8 | | -0.1 | | -0.1 | |
| | | | RMSD | 2 | 0.86 | 0.8 | 0.35 | 0.81 | 0.35 | 1.52 | 0.79 | 0.56 | 0.25 | 0.58 | 0.25 |
| | | | $R^2$ | 0.976 | 0.976 | 0.996 | 0.996 | 0.996 | 0.996 | 0.98 | 0.98 | 0.998 | 0.998 | 0.998 | 0.998 |

| **Opt:** | **b3lyp/6-311+g(d,p)** | | **b3lyp/6-311+g(d,p)** | | **b3lyp/6-311+g(d,p)** | | **mpw1p/w916-31(d,p)** | | **b3lyp/6-311+g(d,p)** | | **b3lyp/6-311+g(d,p)** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **J-calc** | **b3lyp/6-31g(d)** | | **b3lyp/6-311g(d,p)** | | **b3lyp/6-311+g(d,p)** | | **mpw1pw91/6-31g(d)** | | **mpw1pw91/6-311g(d,p)** | | **mpw1pw91/6-311+g(d,p)** | |
| | 6 (1.4) | 7.2 (0.2) | 6.7 (0.8) | 7.4 (0) | 6.6 (0.8) | 7.4 (0) | 6.4 (1) | 7.4 (0.1) | 6.8 (0.6) | 7.5 (0.1) | 6.8 (0.7) | 7.5 (0.1) |
| | 6.2 (1.7) | 7.4 (0.4) | 6.9 (0.9) | 7.7 (0.2) | 6.9 (0) | 7.7 (0.2) | 6.6 (1.2) | 7.6 (0.2) | 7.1 (0) | 7.8 (0.1) | 7.1 (0) | 7.8 (0.1) |
| | 7.2 (1.7) | 8.9 (0) | 8 (0.9) | 8.9 (0) | 8 (0) | 8.9 (0.1) | 7.6 (1.3) | 8.9 (0) | 8.1 (0) | 8.9 (0) | 8.1 (0) | 8.9 (0.1) |
| | 6.2 (1.5) | 7.5 (0.2) | 7 (0.7) | 7.8 (0.1) | 7 (0) | 7.8 (0) | 6.7 (1.1) | 7.6 (0.1) | 7.1 (0) | 7.8 (0.1) | 7.1 (0) | 7.8 (0.1) |
| | 5.5 (0.9) | 6.4 (0) | 5.6 (0.8) | 6.2 (0.2) | 5.6 (0) | 6.1 (0.3) | 5.7 (0.8) | 6.3 (0.1) | 5.5 (0) | 6.1 (0.3) | 5.5 (0) | 6.1 (0.3) |
| | 1.1 (2.8) | -0.2 (1.5) | -0.8 (0.9) | -1.3 (0.4) | -1 (0) | -1.5 (0.2) | 0.4 (2.1) | -0.6 (1.1) | -1.5 (0) | -1.5 (0.2) | -1.7 (0) | -1.7 (0) |
| | -4.3 (2.8) | -8.3 (1.1) | -6.4 (0.8) | -7.7 (0.6) | -6.3 (0) | -7.7 (0.6) | -4.8 (2.3) | -7.5 (0.4) | -7 (0) | -7.4 (0.3) | -6.9 (0) | -7.4 (0.3) |
| | -2.5 (3.8) | -5.5 (0.8) | -5.3 (1) | -6.5 (0.2) | -5.3 (0) | -6.6 (0.3) | -3.1 (3.2) | -5.2 (1.1) | -5.9 (0) | -6.3 (0) | -6 (0) | -6.4 (0.1) |
| | -2.4 (2.4) | -5.5 (0.6) | -4.1 (0.8) | -5.1 (0.2) | -4.1 (0) | -5.1 (0.2) | -3.3 (1.6) | -5.5 (0.6) | -4.7 (0) | -5 (0.1) | -4.7 (0) | -5 (0.1) |
| | 6.8 (1) | 8.3 (0.5) | 7.1 (0.6) | 7.9 (0.2) | 7.1 (0) | 7.9 (0.2) | 6.8 (0.9) | 7.8 (0.1) | 7 (0) | 7.7 (0.1) | 7 (0) | 7.7 (0) |
| | 0.1 (3) | -1.6 (1.2) | -1.8 (1.1) | -2.5 (0.4) | -1.8 (0) | -2.4 (0.5) | -0.5 (2.3) | -1.8 (1) | -2.5 (0) | -2.6 (0.3) | -2.5 (0) | -2.5 (0.4) |
| | -1.5 (3.1) | -4.1 (0.5) | -3.6 (1) | -4.6 (0) | -3.5 (0) | -4.5 (0.1) | -2.2 (2.4) | -4.1 (0.5) | -4.3 (0) | -4.5 (0.1) | -4.2 (0) | -4.4 (0.2) |
| | -1 (2.2) | -3.3 (0.1) | -2.4 (0.8) | -3.1 (0.1) | -2.3 (0) | -3 (0.2) | -1.7 (1.5) | -3.4 (0.2) | -3 (0) | -3.2 (0) | -3 (0) | -3.1 (0.1) |
| | 0.4 (2.3) | -1.2 (0.7) | -1.3 (0.6) | -1.8 (0) | -1.2 (0) | -1.8 (0.1) | -0.4 (1.5) | -1.7 (0.2) | -2.1 (0) | -2.2 (0.3) | -2.1 (0) | -2.1 (0.2) |
| | 5.5 (0.8) | 6.4 (0.1) | 5.9 (0.3) | 6.5 (0.3) | 5.9 (0) | 6.5 (0.3) | 5.5 (0.7) | 6.1 (0.1) | 5.7 (0) | 6.3 (0.1) | 5.7 (0) | 6.4 (0.1) |
| | 0.9 (0.5) | -0.5 (1.8) | 1.1 (0.3) | 0.9 (0.5) | 1.1 (0) | 0.9 (0.5) | 1.1 (0.2) | 0.4 (1) | 1.3 (0) | 1.5 (0.1) | 1.3 (0) | 1.5 (0.1) |
| | 6 (0.5) | 7.2 (1.7) | 4.5 (1) | 4.9 (0.7) | 4.5 (0) | 4.9 (0.6) | 6.9 (1.3) | 7.9 (2.4) | 4.6 (0) | 5.1 (0.4) | 4.6 (0) | 5.1 (0.4) |
| | 4.1 (1.3) | 4.4 (1.1) | 4.7 (0.8) | 5.1 (0.3) | 4.7 (0) | 5.1 (0.4) | 4.2 (1.3) | 4.4 (1.1) | 4.7 (0) | 5.2 (0.3) | 4.7 (0) | 5.2 (0.3) |
| | 5.8 (1.2) | 6.9 (0.2) | 6.3 (0.7) | 7 (0.1) | 6.2 (0) | 6.9 (0.1) | 6 (1.1) | 6.7 (0.3) | 6.2 (0) | 6.8 (0.2) | 6.1 (0) | 6.8 (0.3) |
| | 4.4 (0.2) | 4.7 (0.1) | 4.7 (0) | 5.1 (0.4) | 4.7 (0) | 5.1 (0.5) | 4.4 (0.2) | 4.7 (0.1) | 4.5 (0) | 5 (0.3) | 4.5 (0) | 5 (0.4) |
| | 7 (1.5) | 8.6 (0.1) | 7.9 (0.6) | 8.8 (0.3) | 7.9 (0) | 8.8 (0.3) | 7.2 (1.3) | 8.4 (0.1) | 8.2 (0) | 9 (0.5) | 8.2 (0) | 9 (0.5) |
| | -2 (1.8) | -4.8 (1) | -2.9 (0.9) | -3.7 (0) | -2.9 (0) | -3.7 (0.1) | -3 (0.8) | -5 (1.3) | -3.7 (0) | -3.8 (0.1) | -3.6 (0) | -3.8 (0) |
| | -1.3 (2.1) | -3.8 (0.4) | -2 (1.3) | -2.7 (0.6) | -2.1 (0) | -2.8 (0.6) | -2.5 (0.9) | -4.4 (1) | -2.9 (0) | -3.1 (0.3) | -3 (0) | -3.1 (0.3) |
| a | 0.7 | | 0.9 | | 0.9 | | 0.8 | | 0.9 | | 0.9 | |
| b | 1.2 | | 0.3 | | 0.3 | | 0.9 | | -0.1 | | -0.1 | |
| RMSD | 1.98 | 0.84 | 0.81 | 0.33 | 0.82 | 0.33 | 1.51 | 0.81 | 0.55 | 0.23 | 0.57 | 0.24 |
| $R^2$ | 0.977 | 0.977 | 0.996 | 0.996 | 0.996 | 0.996 | 0.979 | 0.979 | 0.998 | 0.998 | 0.998 | 0.998 |

**S3 HMBC**

| **Opt:** | **mpw1pw91/6-311+g(d,p)** | | **mpw1pw91/6-311+g(d,p)** | | **mpw1pw91/6-311+g(d,p)** | | **b3lyp6-3/11+g(d,p)** | | **mpw1pw91/6-311+g(d,p)** | | **mpw1pw91/6-311+g(d,p)** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **J-calc** | **b3lyp/6-31g(d)** | | **b3lyp/6-311g(d,p)** | | **b3lyp/6-311+g(d,p)** | | **mpw1pw91/6-31g(d)** | | **mpw1pw91/6-311g(d,p)** | | **mpw1pw91/6-311+g(d,p)** | |
| | 6 (1.4) | 7.1 (0.3) | 6.7 (0.8) | 7.3 (0.1) | 6.6 (0.8) | 7.3 (0.1) | 6.5 (0.9) | 7.4 (0) | 6.8 (0.7) | 7.4 (0) | 6.7 (0.7) | 7.4 (0) |
| | 6.2 (1.7) | 7.4 (0.5) | 6.9 (0.9) | 7.7 (0.2) | 6.9 (1) | 7.7 (0.2) | 6.7 (1.2) | 7.7 (0.2) | 7.1 (0.8) | 7.7 (0.1) | 7 (0.8) | 7.8 (0.1) |
| | 7.2 (1.7) | 8.8 (0) | 8 (0.9) | 8.9 (0) | 8 (0.9) | 8.9 (0) | 7.6 (1.2) | 8.9 (0.1) | 8.1 (0.8) | 8.9 (0) | 8.1 (0.8) | 8.9 (0) |
| | 6.2 (1.5) | 7.4 (0.3) | 7 (0.7) | 7.8 (0) | 7 (0.7) | 7.7 (0) | 6.7 (1) | 7.7 (0) | 7.1 (0.6) | 7.8 (0.1) | 7.1 (0.7) | 7.8 (0.1) |
| | 5.7 (0.7) | 6.6 (0.2) | 5.8 (0.6) | 6.4 (0.1) | 5.7 (0.7) | 6.3 (0.1) | 5.5 (0.9) | 6.1 (0.3) | 5.7 (0.7) | 6.3 (0.1) | 5.6 (0.8) | 6.2 (0.2) |
| | 1.2 (2.9) | 0 (1.6) | -0.8 (0.9) | -1.3 (0.4) | -1 (0.7) | -1.5 (0.2) | 0.3 (2) | -0.7 (1) | -1.5 (0.2) | -1.5 (0.2) | -1.7 (0) | -1.6 (0) |
| | -4.3 (2.9) | -8.2 (1.1) | -6.4 (0.8) | -7.6 (0.5) | -6.3 (0.8) | -7.6 (0.5) | -4.9 (2.2) | -7.6 (0.4) | -7 (0.2) | -7.4 (0.2) | -6.9 (0.2) | -7.3 (0.2) |
| | -2.4 (3.9) | -5.4 (0.9) | -5.3 (1) | -6.4 (0.1) | -5.3 (0.9) | -6.5 (0.2) | -3.2 (3.1) | -5.3 (1) | -5.9 (0.4) | -6.3 (0) | -6 (0.3) | -6.3 (0.1) |
| | -2.5 (2.3) | -5.6 (0.7) | -4.1 (0.7) | -5.1 (0.2) | -4.1 (0.7) | -5.1 (0.2) | -3.2 (1.6) | -5.4 (0.5) | -4.7 (0.1) | -5 (0.1) | -4.7 (0.1) | -5 (0.1) |
| | 6.8 (1) | 8.2 (0.5) | 7.1 (0.6) | 7.8 (0.1) | 7.1 (0.6) | 7.9 (0.2) | 6.8 (0.9) | 7.8 (0.1) | 7 (0.8) | 7.6 (0.1) | 7 (0.8) | 7.7 (0) |
| | 0.1 (3) | -1.7 (1.2) | -1.9 (1) | -2.5 (0.4) | -1.9 (1) | -2.5 (0.4) | -0.6 (2.3) | -1.8 (1.1) | -2.5 (0.3) | -2.6 (0.3) | -2.5 (0.4) | -2.6 (0.3) |
| | -1.6 (3.1) | -4.1 (0.5) | -3.7 (0.9) | -4.6 (0) | -3.6 (1) | -4.5 (0.1) | -2.2 (2.4) | -4 (0.6) | -4.4 (0.3) | -4.6 (0.1) | -4.3 (0.3) | -4.5 (0.1) |
| | -1 (2.2) | -3.3 (0.1) | -2.4 (0.8) | -3.1 (0.1) | -2.4 (0.8) | -3.1 (0.1) | -1.7 (1.5) | -3.4 (0.2) | -3.1 (0.1) | -3.2 (0) | -3.1 (0.1) | -3.1 (0.1) |
| | 0.4 (2.3) | -1.2 (0.6) | -1.4 (0.5) | -1.9 (0) | -1.3 (0.5) | -1.8 (0) | -0.4 (1.4) | -1.7 (0.2) | -2.2 (0.3) | -2.2 (0.3) | -2.1 (0.3) | -2.2 (0.3) |
| | 5.6 (0.7) | 6.5 (0.2) | 6 (0.2) | 6.6 (0.3) | 6 (0.2) | 6.6 (0.4) | 5.4 (0.8) | 6 (0.2) | 5.8 (0.4) | 6.4 (0.2) | 5.8 (0.4) | 6.4 (0.2) |
| | 0.9 (0.4) | -0.4 (1.8) | 1.1 (0.2) | 1 (0.4) | 1.1 (0.2) | 1 (0.4) | 1.1 (0.3) | 0.3 (1) | 1.3 (0) | 1.6 (0.2) | 1.3 (0) | 1.6 (0.2) |
| | 6.1 (0.5) | 7.2 (1.7) | 4.5 (1) | 4.9 (0.6) | 4.5 (1) | 4.9 (0.6) | 6.8 (1.3) | 7.9 (2.4) | 4.6 (0.9) | 5.1 (0.4) | 4.6 (1) | 5.1 (0.4) |
| | 4.2 (1.3) | 4.4 (1.1) | 4.8 (0.7) | 5.2 (0.3) | 4.7 (0.8) | 5.1 (0.3) | 4.2 (1.3) | 4.4 (1.1) | 4.7 (0.8) | 5.2 (0.2) | 4.7 (0.8) | 5.2 (0.3) |
| | 6 (1.1) | 7.1 (0) | 6.5 (0.6) | 7.1 (0.1) | 6.4 (0.7) | 7.1 (0) | 5.8 (1.2) | 6.6 (0.5) | 6.3 (0.7) | 7 (0.1) | 6.2 (0.8) | 6.9 (0.2) |
| | 4.5 (0.1) | 4.9 (0.2) | 4.8 (0.2) | 5.2 (0.6) | 4.8 (0.2) | 5.3 (0.6) | 4.3 (0.3) | 4.6 (0) | 4.6 (0.1) | 5.1 (0.4) | 4.6 (0) | 5.1 (0.5) |
| | 6.8 (1.7) | 8.3 (0.2) | 7.7 (0.8) | 8.6 (0) | 7.7 (0.8) | 8.6 (0) | 7.4 (1.1) | 8.7 (0.2) | 8 (0.6) | 8.7 (0.2) | 8 (0.6) | 8.7 (0.2) |
| | -2.2 (1.6) | -5 (1.3) | -3.1 (0.7) | -3.9 (0.1) | -3.1 (0.7) | -3.9 (0.1) | -2.8 (1) | -4.8 (1) | -3.8 (0) | -4 (0.2) | -3.8 (0) | -3.9 (0.2) |
| | -1.3 (2.1) | -3.7 (0.4) | -2.1 (1.2) | -2.8 (0.6) | -2.2 (1.2) | -2.8 (0.6) | -2.6 (0.8) | -4.5 (1.1) | -3 (0.4) | -3.1 (0.3) | -3 (0.4) | -3.1 (0.3) |
| a | 0.7 | | 0.9 | | 0.9 | | 0.8 | | 0.9 | | 0.9 | |
| b | 1.2 | | 0.3 | | 0.3 | | 0.8 | | -0.1 | | -0.1 | |
| RMSD | 1.99 | 0.86 | 0.77 | 0.31 | 0.78 | 0.31 | 1.5 | 0.79 | 0.53 | 0.22 | 0.54 | 0.22 |
| $R^2$ | 0.976 | 0.976 | 0.997 | 0.997 | 0.997 | 0.997 | 0.98 | 0.98 | 0.998 | 0.998 | 0.998 | 0.998 |

## A6.3. Homonuclear coupling constant DFT

**Table A6-3. Comparison of experimental and theoretical *J*-coupling constants calculated as given from structures optimized to B3LYP/6-31G(d). For each dataset the unscaled is given in the left column and a scaled set in the right.**

| H1 | H2 | Exp | B3LYP/6-31G(d,p) fconly | | B3LYP/6-31G(d,p) | | B3LYP/6-311+G(d,p) fconly | | B3LYP/6-311+G(d,p) | | B3LYP/cc-pVDZ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 7.7 | 8.3 (0.7) | 7.6 (0.1) | 8.2 (0.5) | 7.5 (0.2) | 9 (1.3) | 8.2 (0.5) | 8.6 (1) | 7.9 (0.2) | 8.2 (0.6) | 7.5 (0.2) |
| 1 | 3 | 1.3 | 1.2 (0) | 1.1 (0.2) | 1.1 (0.2) | 1 (0.3) | 1.2 (0.1) | 1.1 (0.2) | 0.9 (0.3) | 0.9 (0.4) | 0.6 (0.7) | 0.6 (0.7) |
| 3 | 1 | 1.5 | 1.2 (0.3) | 1.1 (0.4) | 1.1 (0.4) | 1 (0.5) | 1.2 (0.3) | 1.1 (0.4) | 0.9 (0.6) | 0.9 (0.6) | 0.6 (0.9) | 0.6 (0.9) |
| 1 | 4 | 0.7 | 0.6 (0.1) | 0.5 (0.1) | 0.6 (0.1) | 0.5 (0.1) | 0.6 (0.1) | 0.5 (0.1) | 0.4 (0.2) | 0.4 (0.3) | 0.2 (0.5) | 0.1 (0.5) |
| 4 | 1 | 0.7 | 0.6 (0) | 0.5 (0.1) | 0.6 (0.1) | 0.5 (0.1) | 0.6 (0.1) | 0.5 (0.1) | 0.4 (0.2) | 0.4 (0.2) | 0.2 (0.5) | 0.1 (0.5) |
| 2 | 4 | 1.2 | 1.1 (0.2) | 1 (0.3) | 0.9 (0.3) | 0.9 (0.4) | 1 (0.2) | 0.9 (0.3) | 0.7 (0.5) | 0.7 (0.6) | 0.5 (0.8) | 0.4 (0.8) |
| 4 | 2 | 1.1 | 1.1 (0) | 1 (0.1) | 0.9 (0.1) | 0.9 (0.2) | 1 (0.1) | 0.9 (0.1) | 0.7 (0.3) | 0.7 (0.4) | 0.5 (0.6) | 0.4 (0.6) |
| 2 | 3 | 7.6 | 8.2 (0.6) | 7.5 (0.1) | 8 (0.4) | 7.3 (0.3) | 8.9 (1.3) | 8.1 (0.5) | 8.4 (0.9) | 7.7 (0.1) | 8 (0.5) | 7.3 (0.2) |
| 3 | 4 | 8.2 | 9.1 (0.9) | 8.3 (0.1) | 9 (0.8) | 8.2 (0) | 9.7 (1.6) | 8.9 (0.7) | 9.3 (1.2) | 8.5 (0.3) | 9 (0.8) | 8.2 (0) |
| 8 | 12 | 0.0 | -0.1 (0.1) | -0.1 (0.1) | -0.2 (0.2) | -0.2 (0.2) | -0.1 (0.1) | -0.1 (0.1) | -0.3 (0.3) | -0.3 (0.3) | -0.5 (0.5) | -0.5 (0.4) |
| 12 | 8 | -0.2 | -0.1 (0.1) | -0.1 (0.1) | -0.2 (0) | -0.2 (0) | -0.1 (0.1) | -0.1 (0.1) | -0.3 (0.1) | -0.3 (0.1) | -0.5 (0.3) | -0.5 (0.3) |
| 8 | 13 | 10.5 | 10.9 (0.4) | 10 (0.6) | 10.9 (0.4) | 9.9 (0.6) | 11.1 (0.6) | 10.1 (0.4) | 10.6 (0.1) | 9.7 (0.8) | 10.2 (0.4) | 9.3 (1.3) |
| 13 | 8 | 10.6 | 10.9 (0.3) | 10 (0.6) | 10.9 (0.3) | 9.9 (0.6) | 11.1 (0.5) | 10.1 (0.4) | 10.6 (0.1) | 9.7 (0.9) | 10.2 (0.4) | 9.3 (1.3) |
| 8 | 16 | -0.2 | -0.3 (0.1) | -0.3 (0.1) | -0.3 (0.1) | -0.3 (0.1) | -0.4 (0.2) | -0.3 (0.1) | -0.5 (0.3) | -0.5 (0.3) | -0.7 (0.5) | -0.6 (0.4) |
| 16 | 8 | -0.3 | -0.3 (0) | -0.3 (0.1) | -0.3 (0) | -0.3 (0) | -0.4 (0) | -0.3 (0) | -0.5 (0.2) | -0.5 (0.2) | -0.7 (0.4) | -0.6 (0.3) |
| 11a | 12 | 8.5 | 9.2 (0.7) | 8.4 (0.1) | 9.5 (0.9) | 8.6 (0.1) | 9.6 (1.1) | 8.8 (0.2) | 9.7 (1.2) | 8.9 (0.3) | 9.8 (1.3) | 9 (0.4) |
| 11a | 13 | -0.5 | -0.3 (0.1) | -0.3 (0.1) | -0.4 (0.1) | -0.3 (0.1) | -0.3 (0.1) | -0.3 (0.1) | -0.4 (0.1) | -0.3 (0.1) | -0.3 (0.1) | -0.3 (0.2) |
| 11b | 12 | 3.4 | 4.1 (0.8) | 3.8 (0.4) | 4 (0.7) | 3.7 (0.3) | 3.9 (0.6) | 3.6 (0.2) | 3.5 (0.2) | 3.2 (0.2) | 3.4 (0.1) | 3.1 (0.2) |
| 12 | 13 | 3.2 | 3.9 (0.7) | 3.6 (0.3) | 4.2 (1) | 3.8 (0.6) | 4.2 (1) | 3.8 (0.6) | 4.5 (1.3) | 4.1 (0.9) | 4.7 (1.5) | 4.3 (1.1) |
| 12 | 14 | -0.2 | -0.2 (0) | -0.2 (0) | -0.1 (0.1) | -0.1 (0.1) | -0.2 (0) | -0.2 (0) | 0.1 (0.3) | 0.1 (0.3) | 0.4 (0.6) | 0.4 (0.6) |
| 12 | 15a | -0.1 | -0.1 (0) | -0.1 (0) | -0.1 (0) | -0.1 (0) | -0.1 (0) | -0.1 (0) | -0.2 (0) | -0.2 (0) | -0.2 (0.1) | -0.2 (0.1) |
| 12 | 23a | -0.1 | -0.1 (0) | -0.1 (0) | -0.2 (0.1) | -0.2 (0.1) | -0.1 (0) | -0.1 (0) | -0.3 (0.2) | -0.3 (0.2) | -0.4 (0.3) | -0.4 (0.3) |
| 12 | 23b | -0.2 | 0 (0.1) | 0 (0.1) | 0.2 (0.4) | 0.2 (0.4) | -0.1 (0) | -0.1 (0) | 0.4 (0.5) | 0.3 (0.5) | 0.8 (1) | 0.7 (0.9) |
| 13 | 14 | 3.1 | 3.6 (0.5) | 3.3 (0.2) | 3.8 (0.7) | 3.4 (0.3) | 3.8 (0.7) | 3.4 (0.3) | 4.1 (1) | 3.7 (0.6) | 4.2 (1) | 3.8 (0.7) |
| 13 | 15a | 0.3 | 0.4 (0.1) | 0.4 (0.1) | 0.4 (0.1) | 0.4 (0.1) | 0.5 (0.2) | 0.5 (0.2) | 0.4 (0.1) | 0.3 (0.1) | 0.3 (0) | 0.2 (0) |
| 13 | 15b | -0.4 | -0.3 (0.1) | -0.3 (0.1) | 0 (0.4) | 0 (0.4) | -0.3 (0.1) | -0.3 (0.1) | 0.3 (0.7) | 0.3 (0.7) | 0.8 (1.2) | 0.7 (1.1) |
| 14 | 15a | 4.9 | 5 (0.1) | 4.6 (0.3) | 5.1 (0.2) | 4.7 (0.3) | 5.2 (0.3) | 4.8 (0.2) | 5.3 (0.4) | 4.9 (0.1) | 5.4 (0.4) | 4.9 (0.1) |
| 14 | 15b | 1.9 | 2.1 (0.2) | 1.9 (0) | 2.2 (0.2) | 2 (0.1) | 2.2 (0.3) | 2 (0.1) | 2.3 (0.4) | 2.1 (0.2) | 2.3 (0.4) | 2.1 (0.2) |
| 14 | 16 | 0.7 | 0.6 (0.1) | 0.5 (0.1) | 0.6 (0.1) | 0.5 (0.1) | 0.6 (0.1) | 0.5 (0.1) | 0.4 (0.3) | 0.4 (0.3) | 0.2 (0.4) | 0.2 (0.5) |
| 14 | 20b | -0.5 | -0.6 (0.1) | -0.5 (0) | -0.6 (0.2) | -0.6 (0.1) | -0.6 (0.2) | -0.6 (0.1) | -0.8 (0.4) | -0.8 (0.3) | -1 (0.5) | -0.9 (0.4) |
| 14 | 22 | -2.9 | -3.3 (0.4) | -3 (0.1) | -3.4 (0.5) | -3.1 (0.2) | -3.3 (0.4) | -3 (0.1) | -3.6 (0.7) | -3.2 (0.3) | -3.7 (0.8) | -3.4 (0.5) |
| 15a | 16 | 4.0 | 4.3 (0.3) | 3.9 (0.1) | 4.4 (0.3) | 4 (0) | 4.5 (0.5) | 4.1 (0.1) | 4.6 (0.6) | 4.2 (0.1) | 4.6 (0.6) | 4.2 (0.2) |
| 15b | 16 | 2.2 | 2.2 (0.1) | 2 (0.1) | 2.3 (0.2) | 2.1 (0) | 2.4 (0.3) | 2.2 (0) | 2.5 (0.4) | 2.3 (0.2) | 2.6 (0.4) | 2.4 (0.2) |
| 16 | 17ab | -0.2 | -0.2 (0) | -0.1 (0.1) | -0.1 (0.1) | -0.1 (0.1) | -0.1 (0.1) | -0.1 (0.1) | -0.1 (0.1) | 0 (0.1) | 0 (0.2) | 0 (0.2) |
| 16 | 18a | -0.5 | -0.5 (0) | -0.5 (0) | -0.6 (0.1) | -0.6 (0) | -0.5 (0.1) | -0.4 (0.1) | -0.7 (0.1) | -0.6 (0.1) | -0.7 (0.2) | -0.7 (0.1) |
| 16 | 18b | 0.1 | 0.1 (0) | 0.1 (0) | 0 (0.1) | 0 (0.1) | 0 (0.1) | 0 (0.1) | -0.3 (0.4) | -0.3 (0.4) | -0.5 (0.6) | -0.5 (0.5) |
| 16 | 20b | -0.3 | -0.4 (0.1) | -0.3 (0.1) | -0.4 (0.2) | -0.4 (0.1) | -0.4 (0.1) | -0.3 (0.1) | -0.6 (0.3) | -0.5 (0.3) | -0.8 (0.5) | -0.7 (0.5) |
| 20a | 22 | -1.4 | -2 (0.6) | -1.8 (0.5) | -2.1 (0.7) | -1.9 (0.6) | -2 (0.6) | -1.8 (0.5) | -2.2 (0.8) | -2 (0.6) | -2.1 (0.7) | -1.9 (0.6) |
| 20b | 22 | -0.6 | -0.6 (0) | -0.6 (0) | -0.4 (0.2) | -0.4 (0.2) | -0.6 (0.1) | -0.6 (0) | -0.2 (0.4) | -0.2 (0.4) | 0.2 (0.7) | 0.1 (0.7) |
| 22 | 23a | 7.0 | 7.5 (0.6) | 6.9 (0.1) | 7.8 (0.8) | 7.1 (0.1) | 8.1 (1.1) | 7.4 (0.4) | 8.3 (1.3) | 7.5 (0.6) | 8.1 (1.1) | 7.4 (0.4) |
| 22 | 23b | 6.1 | 6.7 (0.6) | 6.1 (0) | 6.5 (0.4) | 6 (0.2) | 6.5 (0.4) | 6 (0.2) | 6.1 (0) | 5.6 (0.5) | 5.9 (0.3) | 5.3 (0.8) |
| | | RMSD | 0.36 | 0.21 | 0.40 | 0.26 | 0.55 | 0.27 | 0.57 | 0.41 | 0.66 | 0.59 |
| | | $R^2$ | 0.997 | 0.997 | 0.994 | 0.994 | 0.994 | 0.994 | 0.986 | 0.986 | 0.971 | 0.971 |

## A6.4. Theoretical basis of the $S^3$ HMBC hetero pulse sequence

Theoretically there are two pure zero-quantum operators, $2\hat{I}_y\hat{S}_y+2\hat{I}_x\hat{S}_x$ and $2\hat{I}_y\hat{S}_x-2\hat{I}_x\hat{S}_y$, and the operator $2\hat{I}_y\hat{S}_y+2\hat{I}_x\hat{S}_x$ was chosen.[27,28] This needed to be translated into a pulse sequence which is achieved in the following equations, starting from the $ZQ_x$ operator. For ease the operator(s) which are used in the new equation is given in bold. The mathematical functions, most of them very simple, in the equations (6.2) to (6.7) are used throughout and thus included in the box below.[28]

$$e^{A+B} = e^A e^B \tag{6.2}$$

$$e^{i\varphi B} A e^{-i\varphi B} = A\cos\varphi - i[B,A]\sin\varphi \tag{6.3}$$

$$\cos\frac{\pi}{2} = 0 \tag{6.4}$$

$$\sin\frac{\pi}{2} = 0 \tag{6.5}$$

$$e^{i\varphi B} e^{-i\varphi B} = 1 \tag{6.6}$$

$$U^{-1}e^{i\mathcal{H}\tau}U = U^{-1}\left\{\sum_{k=0}^{\infty}\frac{1}{k!}(i\mathcal{H}\tau)^k\right\}U = 1 + i\tau U^{-1}\mathcal{H}U + \frac{(i\tau)^2}{2!}(U^{-1}\mathcal{H}U)(U^{-1}\mathcal{H}U)\ldots$$
$$= \sum_{k=0}^{\infty}\frac{1}{k!}(i\tau U^{-1}\mathcal{H}U)^k = e^{iU^{-1}\mathcal{H}U\tau} \tag{6.7}$$

Starting from the zero quantum operator, and using (6.2):

$$ZQ_x = 2\hat{I}_y\hat{S}_y + 2\hat{I}_x\hat{S}_x \tag{6.8}$$

$$e^{-i\pi ZQ_x} = e^{-i\pi(2\hat{I}_y\hat{S}_y+2\hat{I}_x\hat{S}_x)} = \boldsymbol{e^{-i\pi(2\hat{I}_y\hat{S}_y)}}e^{-i\pi(2\hat{I}_x\hat{S}_x)} \tag{6.9}$$

The two operators are expanded in isolation, starting with $e^{-i\pi(2\hat{I}_y\hat{S}_y)}$, which is expanded by adding operators that equal 1 to both sides of the operator, (6.6).

$$e^{-i\pi(2\hat{I}_y\hat{S}_y)} = e^{-i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}\boldsymbol{e^{i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}}\boldsymbol{e^{-i\pi(2\hat{I}_y\hat{S}_y)}}\boldsymbol{e^{-i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}}e^{i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)} \tag{6.10}$$

The three middle operators are manipulated, using equation (6.7) and (6.2).

$$e^{i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}e^{-i\pi(2\hat{I}_y\hat{S}_y)}e^{-i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)} = e^{\left[-i\pi e^{i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}(2\hat{I}_y\hat{S}_y)e^{-i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}\right]}$$
$$= e^{\left[-i\pi\boldsymbol{e^{i\frac{\pi}{2}(\hat{I}_x)}}e^{i\frac{\pi}{2}(\hat{S}_x)}(\boldsymbol{2\hat{I}_y\hat{S}_y})\boldsymbol{e^{-i\frac{\pi}{2}(\hat{I}_x)}}e^{-i\frac{\pi}{2}(\hat{S}_x)}\right]} \tag{6.11}$$

Using equation (6.3)

$$e^{i\frac{\pi}{2}(\hat{I}_x)}\left(2\hat{I}_y\hat{S}_y\right)e^{-i\frac{\pi}{2}(\hat{I}_x)} = -i\left[\hat{I}_x, 2\hat{I}_y\hat{S}_y\right] = -ii\left[2\hat{I}_z\hat{S}_y\right] = \left[2\hat{I}_z\hat{S}_y\right] \tag{6.12}$$

Returning the answer to equation (6.11)

$$e^{i\frac{\pi}{2}(\hat{S}_x)}\left(2\hat{I}_z\hat{S}_y\right)e^{-i\frac{\pi}{2}(\hat{S}_x)} = -i\left[\hat{S}_x, 2\hat{I}_z\hat{S}_y\right] = -ii\left[2\hat{I}_z\hat{S}_z\right] = \left[2\hat{I}_z\hat{S}_z\right] \tag{6.13}$$

Thus

$$e^{\left[-i\pi e^{i\frac{\pi}{2}(\hat{I}_x)}(2\hat{I}_y\hat{S}_y)e^{-i\frac{\pi}{2}(\hat{I}_x)}\right]} = e^{-i\pi(2\hat{I}_z\hat{S}_y)} \quad \textbf{(6.14)}$$

$$e^{\left[-i\pi e^{i\frac{\pi}{2}(\hat{S}_x)}(2\hat{I}_z\hat{S}_y)e^{-i\frac{\pi}{2}(\hat{S}_x)}\right]} = e^{-i\pi(2\hat{I}_z\hat{S}_z)} \quad \textbf{(6.15)}$$

Which leads to (6.11) being equal to

$$e^{-i\pi(2\hat{I}_y\hat{S}_y)} = e^{-i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}e^{-i\pi(2\hat{I}_z\hat{S}_z)}e^{i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)} \quad \textbf{(6.16)}$$

An identical approach may be taken for $e^{-i\pi(2\hat{I}_x\hat{S}_x)}$ giving

$$e^{-i\pi(2\hat{I}_x\hat{S}_x)} = e^{-i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)}\boldsymbol{e^{i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)}e^{-i\pi(2\hat{I}_x\hat{S}_x)}e^{-i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)}}e^{i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)} \quad \textbf{(6.17)}$$

$$\begin{aligned} e^{i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)}e^{-i\pi(2\hat{I}_x\hat{S}_x)}e^{-i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)} &= e^{\left[-i\pi e^{i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)}(2\hat{I}_x\hat{S}_x)e^{-i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)}\right]} \\ &= e^{\left[-i\pi\boldsymbol{e^{i\frac{\pi}{2}(\hat{I}_y)}}e^{i\frac{\pi}{2}(\hat{S}_x)}\boldsymbol{(2\hat{I}_x\hat{S}_x)e^{-i\frac{\pi}{2}(\hat{I}_y)}}e^{-i\frac{\pi}{2}(\hat{S}_y)}\right]} \end{aligned} \quad \textbf{(6.18)}$$

$$e^{\left[-i\pi e^{i\frac{\pi}{2}(\hat{I}_x)}(2\hat{I}_y\hat{S}_y)e^{-i\frac{\pi}{2}(\hat{I}_x)}\right]} = e^{i\pi(2\hat{I}_z\hat{S}_y)} \quad \textbf{(6.19)}$$

$$e^{\left[-i\pi e^{i\frac{\pi}{2}(\hat{S}_y)}(2\hat{I}_z\hat{S}_x)e^{-i\frac{\pi}{2}(\hat{S}_y)}\right]} = e^{-i\pi(2\hat{I}_z\hat{S}_z)} \quad \textbf{(6.20)}$$

$$e^{-i\pi(2\hat{I}_x\hat{S}_x)} = e^{-i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)}e^{-i\pi(2\hat{I}_z\hat{S}_z)}e^{i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)} \quad \textbf{(6.21)}$$

This result in equation (6.9) being expanded to

$$e^{-i\pi(2\hat{I}_y\hat{S}_y+2\hat{I}_x\hat{S}_x)} = e^{-i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}e^{-i\pi(2\hat{I}_z\hat{S}_z)}\boldsymbol{e^{i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}e^{-i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)}}e^{-i\pi(2\hat{I}_z\hat{S}_z)}e^{i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)} \quad \textbf{(6.22)}$$

Next we consider the middle two operators, which are expanded by adding operators that equal 1 before the two operators, (6.6).

$$e^{i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}e^{-i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)} = e^{-i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)}\boldsymbol{e^{i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)}e^{i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}e^{-i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)}} \quad \textbf{(6.23)}$$

In an equal approach to the manipulation of equation (6.11), the last part gives

$$e^{i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)}e^{i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}e^{-i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)} = e^{\left[i\frac{\pi}{2}\boldsymbol{e^{i\frac{\pi}{2}(\hat{I}_y)}}e^{i\frac{\pi}{2}(\hat{S}_y)}\boldsymbol{(\hat{I}_x+\hat{S}_x)e^{-i\frac{\pi}{2}(\hat{I}_y)}}e^{-i\frac{\pi}{2}(\hat{S}_y)}\right]} \quad \textbf{(6.24)}$$

$$e^{\left[i\frac{\pi}{2}e^{i\frac{\pi}{2}(\hat{I}_y)}(\hat{I}_x+\hat{S}_x)e^{-i\frac{\pi}{2}(\hat{I}_y)}\right]} = e^{i\frac{\pi}{2}(\hat{I}_z+\hat{S}_x)} \quad \textbf{(6.25)}$$

$$e^{\left[i\frac{\pi}{2}e^{i\frac{\pi}{2}(\hat{S}_y)}(\hat{I}_z+\hat{S}_x)e^{-i\frac{\pi}{2}(\hat{S}_y)}\right]} = e^{i\frac{\pi}{2}(\hat{I}_z+\hat{S}_z)} \quad \textbf{(6.26)}$$

And thus

$$e^{i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}e^{-i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)} = e^{-i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)}e^{i\frac{\pi}{2}(\hat{I}_z+\hat{S}_z)} \quad \textbf{(6.27)}$$

This result in equation (6.9) being expanded to

$$e^{-i\pi(2\hat{I}_y\hat{S}_y+2\hat{I}_x\hat{S}_x)} = e^{-i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}e^{-i\pi(2\hat{I}_z\hat{S}_z)}e^{-i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)}e^{i\frac{\pi}{2}(\hat{I}_z+\hat{S}_z)}e^{-i\pi(2\hat{I}_z\hat{S}_z)}e^{i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)} \quad \textbf{(6.28)}$$

Finally, operators that equal 1 are appended after the last operator of the sequence, (6.6).

$$e^{-i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}e^{-i\pi(2\hat{I}_z\hat{S}_z)}e^{-i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)}e^{i\frac{\pi}{2}(\hat{I}_z+\hat{S}_z)}e^{-i\pi(2\hat{I}_z\hat{S}_z)}e^{i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)} \quad \textbf{(6.29)}$$
$$= e^{-i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}e^{-i\pi(2\hat{I}_z\hat{S}_z)}e^{-i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)}\boldsymbol{e^{i\frac{\pi}{2}(\hat{I}_z+\hat{S}_z)}e^{-i\pi(2\hat{I}_z\hat{S}_z)}e^{i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)}e^{-i\frac{\pi}{2}(\hat{I}_z+\hat{S}_z)}e^{i\frac{\pi}{2}(\hat{I}_z+\hat{S}_z)}}$$

Where the last part is then considered

$$e^{i\frac{\pi}{2}(\hat{I}_z+\hat{S}_z)}e^{-i\pi(2\hat{I}_z\hat{S}_z)}e^{i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)}e^{-i\frac{\pi}{2}(\hat{I}_z+\hat{S}_z)}e^{i\frac{\pi}{2}(\hat{I}_z+\hat{S}_z)} \quad \textbf{(6.30)}$$
$$= e^{-i\pi(2\hat{I}_z\hat{S}_z)}\boldsymbol{e^{i\frac{\pi}{2}(\hat{I}_z+\hat{S}_z)}e^{i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)}e^{-i\frac{\pi}{2}(\hat{I}_z+\hat{S}_z)}}e^{i\frac{\pi}{2}(\hat{I}_z+\hat{S}_z)}$$

In an equal approach to the manipulation of equation (6.11), the middle part gives

$$e^{i\frac{\pi}{2}(\hat{I}_z+\hat{S}_z)}e^{i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)}e^{-i\frac{\pi}{2}(\hat{I}_z+\hat{S}_z)} = e^{\left[i\frac{\pi}{2}\boldsymbol{e^{i\frac{\pi}{2}(\hat{I}_z)}}e^{i\frac{\pi}{2}(\hat{S}_z)}\boldsymbol{(\hat{I}_y+\hat{S}_y)e^{-i\frac{\pi}{2}(\hat{I}_z)}}e^{-i\frac{\pi}{2}(\hat{S}_z)}\right]} \quad \textbf{(6.31)}$$

$$e^{\left[i\frac{\pi}{2}e^{i\frac{\pi}{2}(\hat{I}_z)}(\hat{I}_y+\hat{S}_y)e^{-i\frac{\pi}{2}(\hat{I}_z)}\right]} = e^{i\frac{\pi}{2}(\hat{I}_x+\hat{S}_y)} \quad \textbf{(6.32)}$$

$$e^{\left[i\frac{\pi}{2}e^{i\frac{\pi}{2}(\hat{S}_z)}(\hat{I}_x+\hat{S}_y)e^{-i\frac{\pi}{2}(\hat{S}_z)}\right]} = e^{i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)} \quad \textbf{(6.33)}$$

And thus

$$e^{-i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}e^{-i\pi(2\hat{I}_z\hat{S}_z)}e^{-i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)}e^{i\frac{\pi}{2}(\hat{I}_z+\hat{S}_z)}e^{-i\pi(2\hat{I}_z\hat{S}_z)}e^{i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)}e^{-i\frac{\pi}{2}(\hat{I}_z+\hat{S}_z)}e^{i\frac{\pi}{2}(\hat{I}_z+\hat{S}_z)} \quad \textbf{(6.34)}$$
$$= e^{-i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}e^{-i\pi(2\hat{I}_z\hat{S}_z)}e^{-i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)}e^{-i\pi(2\hat{I}_z\hat{S}_z)}e^{i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}e^{i\frac{\pi}{2}(\hat{I}_z+\hat{S}_z)}$$

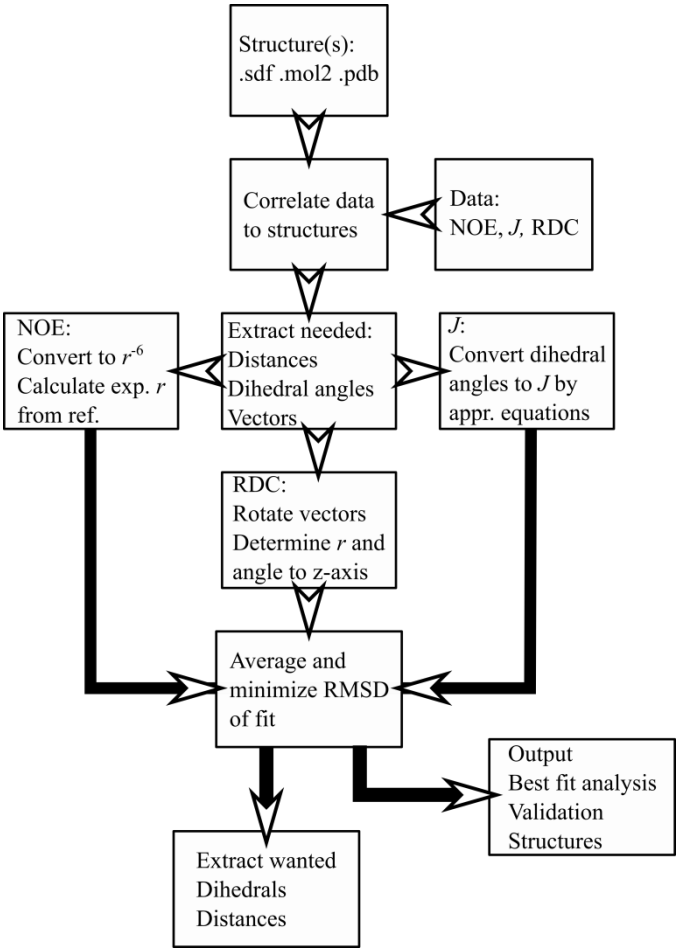
The last z-operator constitutes a phase shift and may be ignored.

Thus the full transformation is finished, as equation (6.9) has been transformed to equation (6.35).

$$e^{-i\pi(2\hat{I}_y\hat{S}_y+2\hat{I}_x\hat{S}_x)} = e^{-i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)}e^{-i\pi(2\hat{I}_z\hat{S}_z)}e^{-i\frac{\pi}{2}(\hat{I}_y+\hat{S}_y)}e^{-i\pi(2\hat{I}_z\hat{S}_z)}e^{i\frac{\pi}{2}(\hat{I}_x+\hat{S}_x)} \quad \textbf{(6.35)}$$

# A7. The DICONO program

The author is not a programmer, nor a software developer. The scripts used were crude and might certainly be optimized. That being said the approach is sturdy, and time was allocated to increase efficiency and decrease the time of calculation for multiple input structures as well as a high degree of structure rotation. The workflow of the program is outlined below. The actual scripts are not included but are provided upon request.




## A7.1. Input

Structures in the input format .pdb, .mol2 and .sdf are accepted, but only the two latter formats work as multistructure-files. The input files need the numbering from the 3D structures, but also include a name column for easier interpretation of data. Examples of the different inputs (NOE, *J*, RDC) are given in Table A7-1.

**Table A7-1. Input format for the NOE (left), *J*-coupling constant (center) and RDC (right) calculations. Upper line is explanation and should not be included in actual scripts. NOE: Upper line is the reference line number which determinates the reference. From left: Names of H1 and H2, number in the 3D structure of H1 and H2, the intensity from PANIC or Macura or other, used (0 use, 1 ignore). *J*-coupling constant: From left: Names of H1 and H2, number in the 3D structure of H1, X1, X2 and H2, the experimental *J*-coupling constant in Hz and the type of coupling (1: NH-CH (Kaplus), 2: CH-CH (Karplus) or 3: CH-CH (HLA)). RDC: From left: Number in the 3D structure of X1, X2, and the experimental RDC in Hz.**

| H1 | H2 | N 1 | N 2 | $\eta$ | Used |
|---|---|---|---|---|---|
| 36 | | | | | |
| 22 | 20b | 45 | 44 | 0.01268 | 0 |
| 22 | 18b | 45 | 42 | 0.00298 | 0 |
| 22 | 12 | 45 | 33 | 0.00363 | 1 |
| 12 | 13 | 33 | 34 | 0.01276 | 0 |

| H1 | H2 | N 1 (H) | N 2 (X) | N 3 (X) | N 4 (H) | *J* | Type |
|---|---|---|---|---|---|---|---|
| 8 | 13 | 30 | 8 | 13 | 34 | 10.5 | 3 |
| 11a | 12 | 32 | 11 | 12 | 33 | 8.3 | 3 |
| 11b | 12 | 31 | 11 | 12 | 33 | 3.2 | 3 |
| 12 | 13 | 33 | 12 | 13 | 34 | 3.2 | 3 |

| N 1 | N 2 | *D* |
|---|---|---|
| 3 | 28 | 23.77 |
| 22 | 45 | 0.74 |
| 2 | 27 | 13.76 |
| 1 | 26 | 26.57 |

### Correlate data to structure

The input data is evaluated and groups need averaging, e.g. methyl groups are identified.

### Extract observables

The distances, dihedrals and vectors are determined from the input data, for RDCs the structures are rotated.

### NOE distances

The distances of the input data are extracted and converted to $r^{-6}$. Distances relative to the reference distance are determined. Note that the reference distance is re-evaluated during the fitting procedure, and need not be a static distance.

### *J*-coupling constants

Dihedrals are extracted and converted to *J*-coupling constants for all structures using either Karplus or the HLA equation. The participating groups and thus constants in the HLA equation are detected automatically if HLA is used.

### RDCs

The implementation is described in Chapter 8 for the $\theta$-method.

### Fitting

The data was fitted by averaging the properties (distances, J-coupling constants, RDCs) of the structures, and by iteratively adding structures that best fit the data from the pool of structures, thus constructing a best fit average. The RMSD of NOEs and *J*-coupling constants and *Q*-factor of RDCs were used for evaluation of fit.

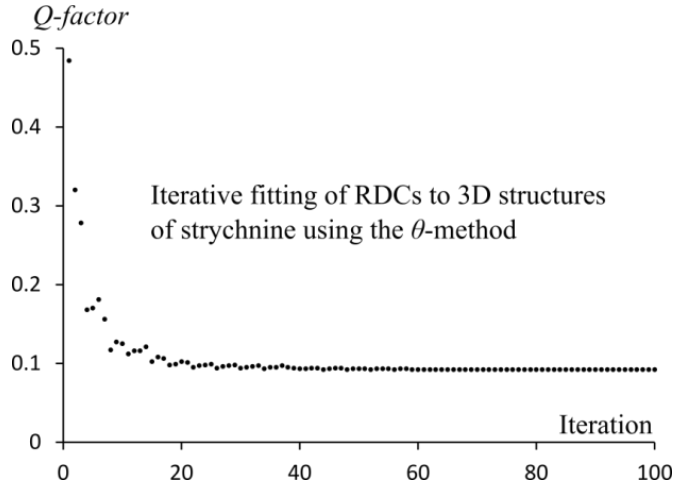
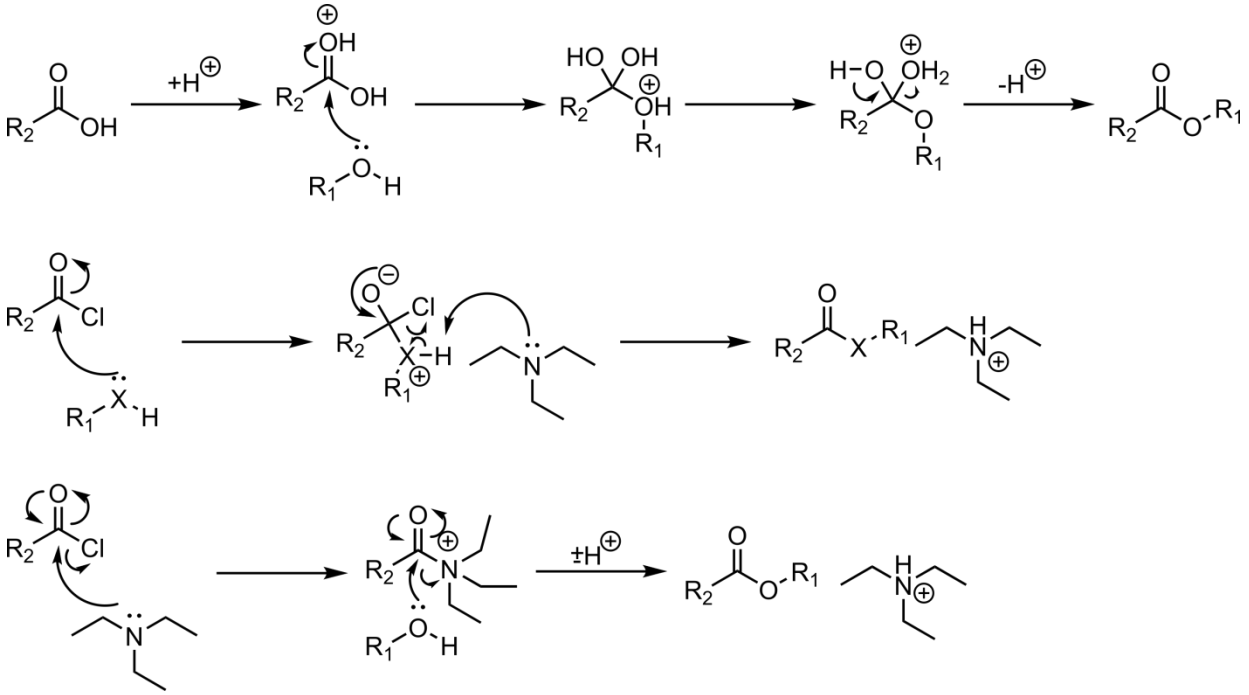



**Figure A7.1. Iterative fitting of strychnine, development of *Q* by addition of structures.**

## A8. Relevant reactions not shown in the thesis

The relevant proposed reaction mechanisms for reactions of an alcohol to an carboxylic acid and reactions with an acid chloride are found below.[29,30] Note that two mechanisms have been proposed when using an alcohol as reactant.



## A9. Correction of $\tau_c$

Below is given the data needed to calculate the dependence of the experimental T-ROE intensity of the offset of the transmitter. Calculated using the formula in Section 3.5 in the thesis. The difference is reduced significantly by correcting the data. If no correction was needed the data at different offsets would be identical.

$\gamma B_1$: 5452 Hz

| δ(H1) | δ(H2) | $\eta_{NOE}$ |
|---|---|---|
| 8.42 | 7.42 | 11.94 |
| 3.43 | 5.46 | 14.61 |
| 3.43 | 2.59 | 87.59 |
| 3.43 | 1.7 | 22.05 |

$\omega_T$: 2498.35 Hz

| $\eta_{T\text{-}ROE}$ | $(\omega_{H1}- \omega_T)$ [Hz] | $(\omega_{H2}- \omega_T)$ [Hz] | $\theta_1$ | $\theta_2$ | $\eta_{ROE}$ |
|---|---|---|---|---|---|
| 31.3 | 1711.65 | 1211.65 | 0.30 | 0.22 | 48.34 |
| 25.25 | -783.35 | 231.65 | -0.14 | 0.04 | 36.27 |
| 195.74 | -783.35 | -1203.35 | -0.14 | -0.22 | 297.69 |
| 41.38 | -783.35 | -1648.35 | -0.14 | -0.29 | 59.42 |

$\omega_T$: 3497.69 Hz

| $\eta_{T\text{-}ROE}$ | $(\omega_{H1}- \omega_T)$ [Hz] | $(\omega_{H2}- \omega_T)$ [Hz] | $\theta_1$ | $\theta_2$ | $\eta_{ROE}$ |
|---|---|---|---|---|---|
| 28.96 | 712.31 | 212.31 | 0.13 | 0.04 | 45.86 |
| 25.80 | -1782.69 | -767.69 | -0.32 | -0.14 | 36.28 |
| 209.91 | -1782.69 | -2202.69 | -0.32 | -0.38 | 306.90 |
| 42.99 | -1782.69 | -2647.69 | -0.32 | -0.45 | 59.10 |

| Corrected | | | Uncorrected | | |
|---|---|---|---|---|---|
| $\eta_{ROE}$(2498 Hz) | $\eta_{ROE}$(3498 Hz) | Diff [%] | $\eta_{ROE}$(2498 Hz) | $\eta_{ROE}$(3498 Hz) | Diff [%] |
| 48.34 | 45.86 | 5.4 | 50.66 | 45.98 | 10.2 |
| 36.27 | 36.28 | 0.0 | 35.89 | 36.99 | -3.0 |
| 297.69 | 306.90 | -3.0 | 303.89 | 332.23 | -8.5 |
| 59.42 | 59.10 | 0.5 | 60.71 | 63.93 | -5.0 |

## A10. References


(1) Gottlieb, H. E.; Kotlyar, V.; Nudelman, A. *J. Org. Chem.* **1997**, *62* (21), 7512–7515.
(2) Enthart, A.; Freudenberger, J. C.; Furrer, J.; Kessler, H.; Luy, B. *J. Magn. Reson.* **2008**, *192*, 314–322.
(3) Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R. *J. Chem. Phys.* **1979**, *71*, 4546–4553.
(4) Kumar, A.; Ernst, R. R.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1980**, *95*, 1–6.
(5) Piotto, M.; Saudek, V.; Sklenar, V. *J. Biomol. NMR* **1992**, *2*, 661–666.
(6) Sklenar, V.; Piotto, M.; Leppik, R.; Saudek, V. *J. Magn. Reson. Ser. A* **1993**, *102*, 241–245.
(7) Hwang, T.-L.; Shaka, A. J. *J. Am. Chem. Soc.* **1992**, *114*, 3157–3159.
(8) Bax, A.; Davis, D. G. *J. Magn. Reson.* **1985**, *63*, 207–213.
(9) Aue, W. P.; Bartholdi, E.; Ernst, R. R. *J. Chem. Phys.* **1976**, *64*, 2229–2246.
(10) Piantini, U.; Sørensen, O. W.; Ernst, R. R. *J. Am. Chem. Soc* **1982**, *104*, 6800–6801.
(11) Hurd, R. E. *J. Magn. Reson.* **1990**, *87* (422-428).
(12) Davis, A. L.; Laue, E. D.; Keeler, J.; Moskau, D.; Lohman, J. *J. Magn. Reson.* **1991**, *94*, 637–644.
(13) Kay, L. E.; Keifer, P.; Saarinen, T. *J. Am. Chem. Soc.* **1992**, *114*, 10663–10665.
(14) A.G. Palmer III; Cavanagh, J.; Wright, P. E.; Rance, M. *J. Magn. Reson.* **1991**, *93*, 151–170.
(15) Willker, W.; Leibfritz, D.; Kerssebaum, R.; Bermel, W. *Magn. Reson. Chem.* **1993**, *31*, 287–292.
(16) Schleucher, J.; Schwendinger, M.; Sattler, M.; Schmidt, P.; Schedletzky, O.; Glaser, S. J.; Sorensen, O. W.; Griesinger, C. *J. Biomol. NMR* **1994**, *4*, 301–306.
(17) Zwahlen, C.; Legault, P.; Vincent, S. J. F.; Greenblatt, J.; Konrat, R.; Kay, L. E. *J. Am. Chem. Soc.* **1997**, *119*, 6711–6721.
(18) Boyer, R. D.; Johnson, R.; Krishnamurthy, K. *J. Magn. Reson.* **2003**, *165*, 253–259.
(19) Cicero, D. O.; Barbato, G.; Bazzo, R. *J. Magn. Reson.* **2001**, *148*, 209–213.
(20) Bax, A.; Summers, M. F. *J. Am. Chem. Soc.* **1986**, *108*, 2093–2094.
(21) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. S.; M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. M.; G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. H.; A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. H.; M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. N.; Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J.; J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. B.; K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. N.; K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. T.; M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. C.; V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. S.; O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. O.; R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. V.; P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. D.; O. Farkas, J. B. Foresman, J. V. Ortiz, J. C.; D. J. Fox. Gaussian, Inc., Wallingford CT 2010.
(22) Maestro. Schrödinger, LLC: New York 2015.
(23) Macromodel. Schrödinger, LLC: New York 2015.
(24) Desmond. D. E. Shaw Research: New York 2015.
(25) Navarro-Vázquez, A. *Magn. Reson. Chem.* **2012**, *50*, S73–S79.
(26) Zweckstetter, M. *Nat. Protoc.* **2008**, *3*, 679–690.
(27) Keeler, J. *Understanding NMR Spectroscopy*, 2nd ed.; John Wiley & Sons, Ltd, 2010.
(28) Sørensen, O. W. *Prog. Nucl. Magn. Reson. Spectrosc.* **1989**, *21*, 503–569.
(29) Clayden, J.; Greeves, N.; Warren, S.; Wothers, P. *Organic Chemistry*; OUP Oxford, 2001.
(30) Hubbard, P.; Brittain, W. J. *J. Org. Chem.* **1998**, *63*, 677–683.

# A11. Publications

**I** Maolanon, Alex R; Villadsen, Jesper S; Christensen, Niels J; **Hoeck, Casper**; Friis, Tina; Harris, Pernille; Gotfredsen, Charlotte H; Fristrup, Peter; Olsen, Christian A. Methyl Effect in Azumamides Provides Insight Into Histone Deacetylase Inhibition by Macrocycles. *Journal of Medicinal Chemistry*, **2014**, *57*, 9644-9657.

**II** Petersen, Lene M; **Hoeck, Casper**; Frisvad, Jens C; Gotfredsen, Charlotte H; Larsen, Thomas O. Dereplication Guided Discovery of Secondary Metabolites of Mixed Biosynthetic Origin from *Aspergillus aculeatus*. *Molecules*, **2014**, *19*, 10898-10921.

**III** Kjaerulff, Louise; Benie, Andrew J; **Hoeck, Casper**; Gotfredsen, Charlotte H; Sørensen, Ole W. $S^3$ HMBC: Spin-State-Selective HMBC for accurate measurement of homonuclear coupling constants. Application to strychnine yielding thirteen hitherto unreported $J_{HH}$. *Journal of Magnetic Resonance*, **2016**, *263*, 101-107

**IV** Li, Ming; **Hoeck, Casper**; Schoffelen, Sanne; Gotfredsen, Charlotte H; Meldal, Morten. Specific Electrostatic Molecular Recognition in Water. *Chemistry - A European Journal*, **2016**, *22*, 7206-7214.

**(V)** **Hoeck, Casper**; Petersen, Lene M; Frisvad, Jens C; Larsen, Thomas O; Gotfredsen, Charlotte H. Homomorphosins A-F, novel diketopiperazines. (Draft)

**(VI)** **Hoeck, Casper**; Gotfredsen, Charlotte H; Sørensen, Ole W. $S^3$ HMBC *hetero*: Spin-State-Selective HMBC for accurate measurement of heteronuclear coupling constants. (Draft)

Journal of Medicinal Chemistry

Article

pubs.acs.org/jmc

# Methyl Effect in Azumamides Provides Insight Into Histone Deacetylase Inhibition by Macrocycles

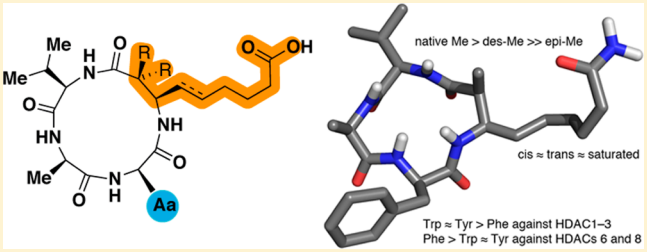
Alex R. Maolanon,[†,#] Jesper S. Villadsen,[†,#] Niels J. Christensen,[†] Casper Hoeck,[†] Tina Friis,[‡] Pernille Harris,[†] Charlotte H. Gotfredsen,[†] Peter Fristrup,[†] and Christian A. Olsen*[,†,§]

[†]Department of Chemistry, Technical University of Denmark, Kemitorvet 207, DK-2800 Kongens Lyngby, Denmark

[‡]Department of Clinical Biochemistry, Immunology and Genetics, Statens Serum Institut, Artillerivej 5, DK-2300 Copenhagen, Denmark

Supporting Information

**ABSTRACT:** Natural, nonribosomal cyclotetrapeptides have traditionally been a rich source of inspiration for design of potent histone deacetylase (HDAC) inhibitors. We recently disclosed the total synthesis and full HDAC profiling of the naturally occurring azumamides (*J. Med. Chem.* **2013**, *56*, 6512). In this work, we investigate the structural requirements for potent HDAC inhibition by macrocyclic peptides using the azumamides along with a series of unnatural analogues obtained through chemical synthesis. By solving solution NMR structures of selected macrocycles and combining these findings with molecular modeling, we pinpoint crucial enzyme−ligand interactions required for potent inhibition of HDAC3. Docking of additional natural products confirmed these features to be generally important. Combined with the structural conservation across HDACs 1−3, this suggests that while cyclotetrapeptides have provided potent and class-selective HDAC inhibitors, it will be challenging to distinguish between the three major class I deacetylases using these chemotypes.




## INTRODUCTION

Histone deacetylase (HDAC) enzymes are a group of epigenetic modulators which catalyze the removal of *ε-N*-acetylated lysine residues found in the N-terminal tails of histone proteins.[1] Furthermore, recent results have shown that HDACs also regulate acetylation levels of a long list of nonhistone proteins[2] and may even target *ε-N*-acyllysine modifications other than acetyl groups.[3−7] Aberrant HDAC expression has been associated with various types of cancer, and these enzymes have therefore been targets in the development of anticancer drugs.[8] Thus far, two HDAC inhibitors have been approved for clinical managing of cutaneous T-cell lymphoma (**1**[9] and **2**[10]), several compounds are in clinical trials for treatment of a variety of additional cancers,[8] and HDAC inhibitors have shown promise in connection with several neurodegenerative disorders,[11] memory,[12] and HIV.[13] Humans have 18 HDAC enzymes, of which 11 are $Zn^{2+}$-dependent (class I: HDACs 1−3 and 8; class IIa: HDACs 4, 5, 7, and 9; class IIb: HDACs 6 and 10; and class IV: HDAC11)[14] and seven are $NAD^{+}$-dependent (class III: sirtuins 1−7).[15] Because of the rich biology affected by these enzymes, there has been a keen interest in trying to develop isoform-selective inhibitors as chemical probes to explore these biological effects in detail and possibly to aid the development of improved therapeutic candidates.

Small molecules exhibiting class selectivity have been discovered, but few truly isoform-selective compounds have surfaced,[16] which may, in part, be due to the high sequence similarity between active sites across the family of $Zn^{2+}$-dependent HDAC enzymes.[14] It has therefore been suggested that interaction with the protein surface surrounding the entrance to the active site, which intrinsically recognizes the acetylated protein substrate, may be favorable for achieving selectivity.[17] The cyclic depsipeptide (e.g., **2**) and cyclic tetrapeptide HDAC inihbitors (e.g., **3**−**5**) have the ability to interact with a significant part of the surface area and due to their macrocyclic nature adopt well-defined three-dimensional display of side chain functionalities. In general, macrocyclic peptides are less susceptible to both endo- and exoproteases than linear peptides, and historically, this class of compounds has delivered several successful drugs such as gramicidin S, cyclosporin A, and polymyxin.[18] Peptide macrocycles have also proven important in the HDAC field as one of the two currently approved drugs is the macrocyclic depsipeptide, romidepsin (**2**).[10] Furthermore, the first isolations of mammalian HDAC enzymes were achieved by using a synthetic analogue of the cyclic tetrapeptide natural product trapoxin (**3**), which enabled affinity matrix pull-down of histone deacetylase proteins from bovine thymus and human Jurkat cell lysate.[19]

Moreover, we found cyclopeptide HDAC inhibitors interesting, because several examples exhibit high inhibitory potencies without containing a strong $Zn^{2+}$-binding group,[20] as for example apicidin (**4**)[21] and the azumamides (**5**).[22−24]

**Received:** September 4, 2014
**Published:** November 7, 2014

   dx.doi.org/10.1021/jm501399d | *J. Med. Chem.* 2014, 57, 9644−9657

Previously, manipulation of the structures of natural cyclotetrapeptides through chemical synthesis have revealed insights into mode of binding[23,25] and structure−activity relationships[16,17] and resulted in novel selectivity profiles.[26,27] For these reasons, we have chosen to explore the azumamide scaffold in detail, which we envisioned could lead to a fundamental understanding of the requirements for potent HDAC inhibition by cyclopeptides and ultimately provide guidelines for future design of macrocyclic HDAC inhibitors.

We recently reported the total syntheses of all five natural azumamides as well as their profiling against the full panel of recombinant human HDACs 1−11.[24] In said study, we observed a detrimental effect on inhibitory potency when changing the stereochemistry of either chiral center in the $\beta$-amino acid residue (gray shading in Figure 1a). Although this drastic effect was not particularly surprising when changing the stereochemistry of the extended $Zn^{2+}$-binding side chain, which must be projected into the active site, the importance of the stereochemistry of the methyl group was less predictable. Because a profound understanding of ligand−enzyme interactions could be of value for future inhibitor design efforts, we decided to address the effect of this methyl group in detail. In addition to the natural products, we therefore prepared versions with desmethyl and dimethyl modification at the $\beta^2$-position in our compound series (**6a**−**9c**, Figure 1b). We also incorporated different naturally occurring residues at the aromatic amino acid position, to include phenyl (**a** series) (trapoxins and azumamides A, D, and E), phenol (**b** series) (azumamides B and C), and indole (**c** series) (apicidins). The unsaturation in the extended side chain was addressed by including a *trans*-olefin version as found in romidepsin (**2**) as well as in another potent macrocyclic HDAC inhibitor largazole[28] (side chain highlighted by orange shading in Figure 1b). Finally, the saturated version mimicking the linkers of trapoxins and apicidins was also included (**7a**−**c**).

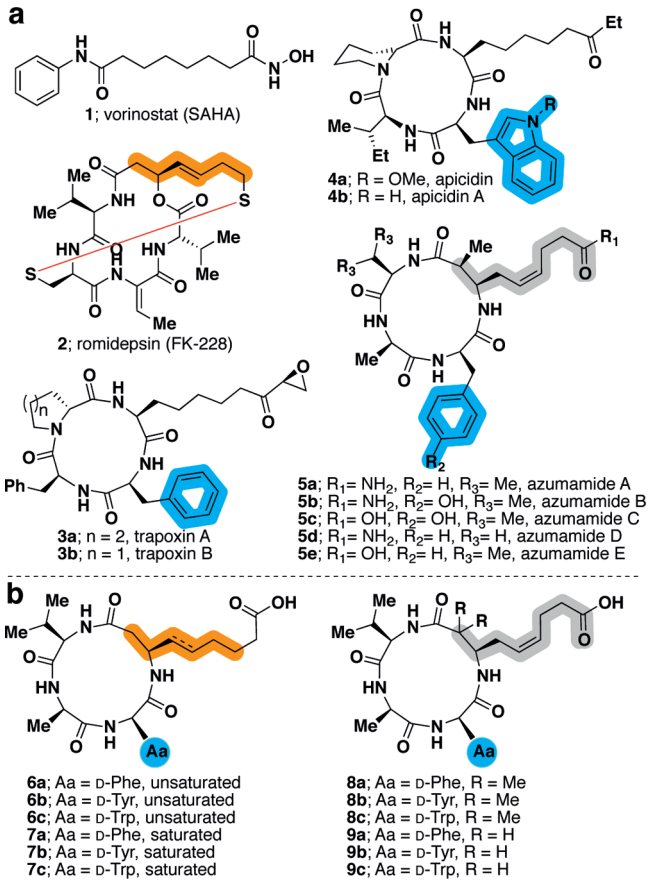



**Figure 1.** Structures of representative HDAC inhibitors (a) and target compounds (b).

## RESULTS

**Chemistry.** Each series of compounds **6a**−**9c** required two key synthetic challenges to be tackled, namely preparation of an appropriate $\beta$-amino acid building block and macrocyclization of linear intermediates to give the final compounds. Preparation of $\beta$-amino acid building blocks for synthesis of **6a**−**c** and **7a**−**c** were prepared from commercially available Boc-L-Asp-O$^t$Bu (**10**) using chemistry adapted from a literature procedure.[29] The $\alpha$-carboxylate of **10** was reduced to the corresponding alcohol and subsequently oxidized under Swern conditions to give an aldehyde, which was immediately subjected to a Wittig reaction to produce alkene **11** (Figure 2a). Installation of the desired side chain by cross-metathesis to give **12** was optimized by investigating a selection of catalysts (Figure 2b). We included a modified Hoveyda−Grubbs catalyst, which contains a nitro substituent on the isopropoxybenzylidene ligand, developed by Grela and co-workers,[30] as this catalyst was previously applied in a transformation similar to ours in a total synthesis of largazole.[31] In addition to the Grela catalyst (Figure 2b, entry 7), we also included the commercially available Zhan's catalyst 1-B (Figure 2b, entry 6), which contains a sulfonamide as electron-withdrawing group in the same position. Comparable yields were obtained using Zhan 1-B, Grubbs second-generation, and Hoveyda−Grubbs second-generation catalysts (Figure 2b), and it was decided to apply 10 mol % Hoveyda−Grubbs second-generation catalyst to afford the desired *trans*-olefin in 67% yield. Acidic deprotection followed by Boc protection of the amino functionality gave the desired building block **13**, and subsequent hydrogenation afforded the second $\beta$-amino acid building block **14** (Figure 2a).

The building block, employed for the synthesis of the dimethylated analogues (**8a**−**c**), was prepared using conditions recently applied in total syntheses of the azumamides.[24] Analogous to the reported protocol, diastereoselective Ellman-type Mannich reaction between **15** (*S*-auxiliary) and **16** afforded the desired *R*-isomer (**17**) in 58% isolated yield (dr = 77:23) as shown in Figure 2c. The absolute stereochemistry was determined by X-ray crystallography upon desilylation with $Bu_4NF$ to give **18**. The alcohol was then elaborated to give the Boc-protected $\beta$-amino acid **20** in 16% overall yield after just four chromatographic steps (Figure 2c). Desmethyl-$\beta$-amino acids **27** (Boc protected) and **28** (Fmoc protected) were prepared by a similar route (Figure 2e), except the (*R*)-auxiliary was required to give the desired stereochemistry (Figure 2d). This selectivity is in agreement with Ellman's original results as well as the proposed Zimmerman−Traxler-type transition state, where coordination to titanium is essential for the high diastereoselectivity. This was further supported by the reversed diastereoselectivity observed when substituting the additive to HMPA, which sequesters lithium ions and thus hampers coordination (Figure 2d, entries 2 and 3). Because of the lower diastereoselectivity combined with difficulties in separation of the *tert*-butyl ester isomers, we chose the conditions shown in entry 1 (Figure 2d) to give **23**, which was then trans-esterified to give the desired *tert*-butyl ester **24** (Figure 2e). Desilylation


dx.doi.org/10.1021/jm501399d | J. Med. Chem. 2014, 57, 9644−9657

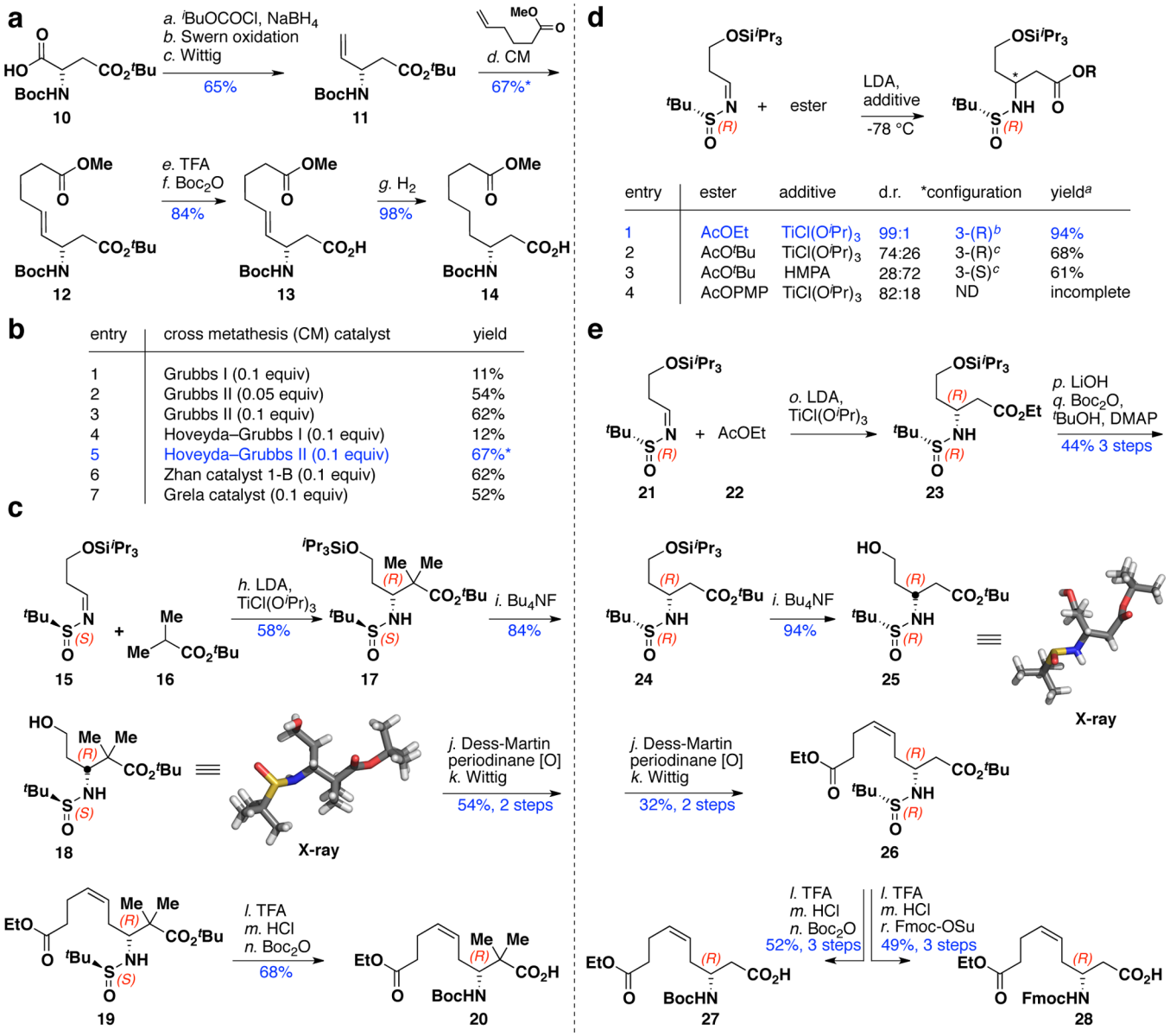



**Figure 2.** Building block syntheses. (a) Preparation of building blocks **13** and **14** from aspartic acid. (b) Optimization of the cross-metathesis reaction to give intermediate **12**. (c) Preparation of dimethylated building block **20** using previously developed Mannich conditions. (d) Optimization of the Mannich reaction to give desmethylated building blocks. [a]Cumulative yield of both diastereomers. [b]Determined by X-ray crystallography upon deprotection and re-esterification as *tert*-butyl ester. [c]Determined by comparison of NMR spectra to those of the compound from entry 1 (Supporting Information, Figure S1). (e) Preparation of desmethylated building blocks **27** and **28**. Reagents and conditions: (*a*) *N*-Methylmorpholine (1 equiv), isobutyl chloroformate (1 equiv), $NaBH_4$ (3 equiv), MeOH, THF, −30 °C → rt, 160 min; (*b*) oxalyl chloride (1.7 equiv), DMSO (3.3 equiv), $Et_3N$ (5 equiv), $CH_2Cl_2$, −78 °C → −40 °C → rt, 2 h; (*c*) $PPh_3CH_3Br$ (2.2 equiv), KHMDS (0.5 M in toluene) (2.1 equiv), THF, −78 °C → rt, 4 h; (*d*) methyl 5-hexenoate (3 equiv), Hoveyda−Grubbs II catalyst (0.1 equiv), $CH_2Cl_2$, 40 °C, 24 h; (*e*) TFA−$CH_2Cl_2$ (1:3), 2 h; (*f*) $Boc_2O$ (1.3 equiv), *i*-$Pr_2NEt$ (2.6 equiv), $CH_2Cl_2$; (*g*) $H_2$, Pd/C (10 wt %, 0.1 equiv), THF, 19 h; (*h*) LDA (2.6 equiv), **16** (2.5 equiv), $TiCl(O^iPr)_3$ (4.5 equiv), THF, −78 °C, 30 min; then **15**, −78 °C, 20 min; (*i*) AcOH (1 equiv), $Bu_4NF$ (2 equiv), THF, 0 °C → rt, 45 min; (*j*) $NaHCO_3$ (1.5 equiv), Dess−Martin periodinane (1.5 equiv), $CH_2Cl_2$, 0 °C → rt, 1 h; (*k*) KHMDS (1.9 equiv), $Ph_3PBr(CH_2)_3CO_2Et$, (2 equiv), THF, −78 °C → rt, 16 h; (*l*) TFA−$CH_2Cl_2$ (1:1, 10 mL, 120−190 equiv), 0 °C → rt, 2 h; (*m*) HCl (4 M in dioxane, 1.8−2.5 equiv), dioxane, 1−3 h; (*n*) *i*-$Pr_2NEt$ (4 equiv), $Boc_2O$ (2 equiv), $CH_2Cl_2$, 20 h; (*o*) LDA (2.1 equiv), **22** (2 equiv), $TiCl(O^iPr)_3$ (4.2 equiv), THF, −78 °C, 30 min, then **21**, −78 °C, 20 min; (*p*) LiOH (4.5 equiv), THF−water, 17 h; (*q*) $Boc_2O$ (1.4 equiv), DMAP (0.3 equiv), *tert*-butanol, 16 h, then $Boc_2O$ (0.3 equiv), 15 h; (*r*) $Na_2CO$ (4 equiv), FmocOSu (1.2 equiv) water−dioxane (5:1, 6 mL), 0 °C → rt, 45 min. ND = not determined.

to give **25** again enabled establishment of the absolute stereochemistry by X-ray crystallography and elaboration of this intermediate provided both the Boc- and Fmoc-protected *β*-amino acids **27** and **28**, respectively (Figure 2e).

All building blocks were incorporated into solid supported linear tetrapeptides, which were cleaved, cyclized using HATU under dilute conditions (0.3−0.8 mM in *N*,*N*-dimethylformamide),[24,27,32] and saponified to give the target compounds **6a**−**9c** in 2−22% isolated yields (Supporting Information, Figure S2). We attribute these yields to two shortcomings in particular; first, the point of cyclization had previously been shown to have an effect on macrolactamization yield in a model system,[24] so synthesis of compound **7c** was attempted by cyclizing between the *β*-amino acid and D-valine (Supporting


dx.doi.org/10.1021/jm501399d | J. Med. Chem. 2014, 57, 9644−9657

Table 1. Inhibition of Recombinant Human Histone Deacetylases (HDACs)[a]

| | $K_i$ ($\mu$M) | | | | | | |
|---|---|---|---|---|---|---|---|
| | class I | | | | class IIb | | class IV |
| compound | HDAC1 | HDAC2 | HDAC3[b] | HDAC8 | HDAC6 | HDAC10 | HDAC11 |
| **6a** | 1.2 ± 0.3 | 1.7 ± 0.04 | 1.4 ± 0.3 | 6 ± 2.7 | I.A.[c] | 1.6 ± 0.03 | 3.2 ± 0.5 |
| **6b** | 0.7 ± 0.2 | 0.6 ± 0.3 | 0.5 ± 0.2 | I.A.[c] | 2.2 ± 0.8 | 0.7 ± 0.03 | 1.3 ± 0.4 |
| **6c** | 0.5 ± 0.06 | 0.4 ± 0.3 | 0.6 ± 0.4 | 3 ± 1.6 | 3.7 ± 0.3 | 0.6 ± 0.06 | 1.1 ± 0.5 |
| **7a** | 0.9 ± 0.4 | 0.8 ± 0.1 | 0.7 ± 0.2 | IA[c] | IA[c] | 1 ± 0.2 | 2 ± 1.1 |
| **7b** | 1 ± 0.6 | 0.3 ± 0.2 | 0.2 ± 0.1 | IA[c] | 3 ± 2 | 0.3 ± 0.2 | 1 ± 0.7 |
| **7c** | 0.9 ± 0.01 | 0.3 ± 0.02 | 0.5 ± 0.2 | 4.3 ± 0.9 | 1.9 ± 0.3 | 0.17 ± 0.03 | 0.3 ± 0.01 |
| **9a** | 0.6 ± 0.2 | 0.9 ± 0.002 | 0.8 ± 0.1 | IA[c] | IA[c] | 0.7 ± 0.003 | 1.5 ± 0.2 |
| **9b** | 0.3 ± 0.2 | 0.3 ± 0.2 | 0.3 ± 0.25 | IA[c] | 6 ± 0.6 | 0.3 ± 0.02 | 0.8 ± 0.3 |
| **9c** | 0.2 ± 0.1 | 0.3 ± 0.2 | 0.5 ± 0.1 | 4.5 ± 0.5 | IA[c] | 0.23 ± 0.09 | 0.9 ± 0.6 |
| SAHA (**1**) | 0.008 | 0.007 | 0.012 | 0.7 | 0.022 | 0.04 | 0.013 |
| apicidin (**4a**)[34] | 0.04 nM | 0.12 nM | 0.26 nM | 0.049 | | | |
| AzuC (**5c**)[24] | 0.032 | 0.04 | 0.014 | >5 | 2 | 0.01 | 0.035 |
| AzuE (**5e**)[24] | 0.067 | 0.05 | 0.025 | 4.4 | >5 | 0.02 | 0.06 |
| epi-AzuE (**36**)[24] | IA | IA | IA | IA | IA | IA | IA |

[a] $IC_{50}$ values were determined from at least two individual dose–response experiments performed in duplicate, and $K_i$ values were calculated based on the Cheng–Prusoff equation. Values are given in $\mu$M unless otherwise noted. [b] In complex with the deacetylase activation domain (DAD) of nuclear receptor corepressor (NCoR1). [c] IA denotes compounds that were "inactive" according to a criterium of exhibiting less than 50% inhibition at 20 $\mu$M. $K_i$ values were not determined for these inhibitor–enzyme combinations due to limited solubility of the ligands at higher concentrations.

Information, Figure S3). Cyclization with the less hindered and more flexible $\beta$-amino acid at the C-terminal afforded a significant increase in isolated yield of **7** (48%, 12 steps), underscoring the importance of point-of-cyclization. Second, we previously demonstrated that purification by reversed-phase high performance liquid chromatography (RP-HPLC) furnished lower yields than column chromatography (normal-phase), albeit at the expense of compound purity,[24] and therefore envision that future syntheses of this class of compounds may benefit from purification by normal-phase preparative HPLC.

**HDAC Profiling.** Initially, two-dose experiments were performed against selected class IIa enzymes (HDAC4 and HDAC7) because the natural products had proven ineffective against this class (Supporting Information, Table S1a). Also, the dimethylated compounds (**8a–c**) were tested against representative enzymes from all classes at two doses to determine whether a full dose–response characterization was warranted (Supporting Information, Table S1b). These experiments showed poor inhibitory activities of all compounds against HDACs 4 and 7 and furthermore showed that the dimethylated compounds were ineffective against HDACs from all classes. Thus, $IC_{50}$ values were determined by full dose–response profiling of compounds **6a–c**, **7a–c**, and **9a–c** on HDACs 1–3, 6, 8, 10, and 11 (Supporting Information, Figure S4). Determination of $K_i$ values was then performed using the Cheng–Prusoff equation [$K_i = IC_{50}/(1 + [S]/K_m)$] under the assumption that all compounds obey a standard fast-on–fast-off mechanism of inhibition as recently demonstrated for azumamide C[33] (Table 1).

In general, the desmethylated analogues exhibited significantly decreased activities against HDACs 1–3, 10, and 11 as compared to the corresponding natural products containing the same aromatic amino acid (**6a**, **7a**, and **9a** vs **5e** as well as **6b**, **7b**, and **9b** vs **5c**). Interestingly, lack of this methyl group did not affect the potencies against HDACs 6 and 8, for which significantly higher $K_i$ values were recorded for both natural products and desmethylated analogues. The greatest difference was observed on HDAC3, where **6a** displayed a 56-fold reduction in activity compared to azumamide E (**5e**). Generally the *trans*-olefin-containing analogues (**6a–c**) exhibited slightly lower potencies (~2-fold) than compounds containing the other two types of extended side chain (i.e., *cis*-olefin or saturated hydrocarbon). The two latter-mentioned side chains, on the other hand, furnished indistinguishable activities. The tyrosine- and tryptophan-containing compounds were ~2–3-fold more potent against HDACs 1–3, 10, and 11 than the corresponding phenylalanine analogues. In contrast, all Tyr-containing compounds failed to inhibit HDAC8 at concentration up to 20 $\mu$M, whereas the Phe- and Trp-containing compounds displayed similar $K_i$ values between 3–6 $\mu$M. However, because all compounds were relatively poor inhibitors of HDAC8 as compared to other class I HDACs (1–3), we decided not to pursue this effect further in the present study.

Overall, the collective HDAC profiling data revealed only minor effects of the choice of aromatic residue or $\beta^3$-side chain identity, generally within an order of magnitude. On the other hand, profiling data emphasized sensitivity of the $\beta^2$-position to modification.

**NMR Spectroscopy.** To rationalize the profiling data and gain a deeper understanding of the structural requirements for potent HDAC inhibition, we solved the solution structures of selected compounds by nuclear magnetic resonance (NMR) spectroscopy combined with molecular modeling. As minute changes in side chain topology are well-known to affect conformational space of small cyclic peptides,[35,36] we suspected that the dramatic changes in biochemical activity might be a result of remodeling the cyclic core of these molecules. Solving the 3D structures by NMR spectroscopy should allow us to


dx.doi.org/10.1021/jm501399d | J. Med. Chem. 2014, 57, 9644–9657

determine whether such conformational changes were responsible for the significant drop in potency observed when altering the substitution pattern or stereochemistry of the 2-position in the $\beta$-amino acid residue.

The compounds chosen for investigation by NMR were azumamide A (**5a**), desmethylated azumamide C (**9b**), and the $\beta^2$-epimer of azumamide E (**36**; Figure 3a), which was

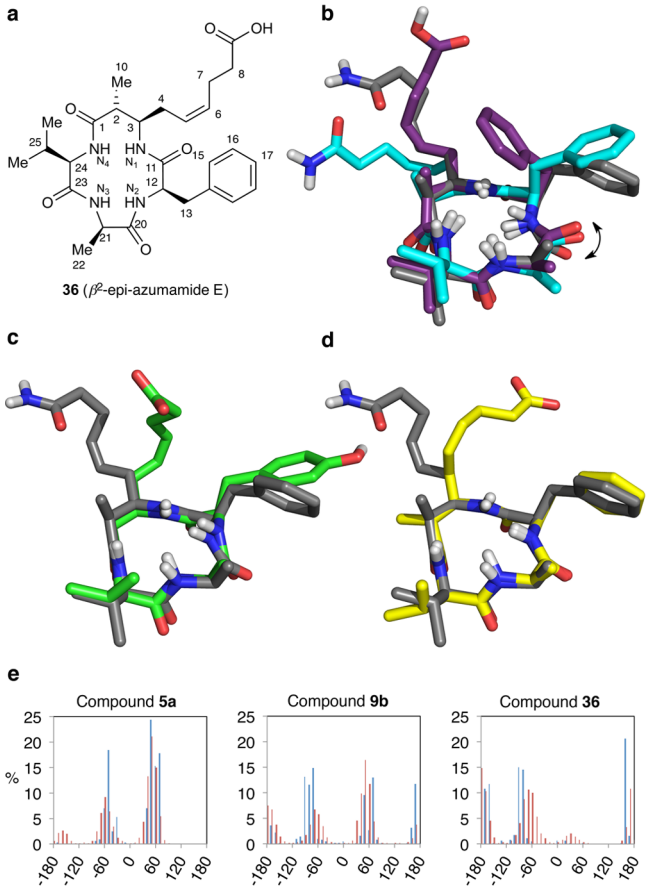



**Figure 3.** NMR solution structures and dihedral angles. (a) Chemical structure of compound **36** with numbering used for NMR assignments. (b) Overlay of azumamide E (**5e**) from the literature (purple),[37] azumamide A (**5a**) from unrestricted simulation (gray), and azumamide A from NOE-restricted simulation (cyan); RMSD = 0.358 Å for the reported NMR structure and our unrestricted azumamide A structure. (c) Overlay of azumamide A (**5a**; gray) with desmethyl compound **9b** (green), RMSD = 0.318 Å. (d) Overlay of azumamide A (**5a**; gray) with $\beta^2$-epi-azumamide E (**36**; yellow), RMSD = 0.324 Å. (e) Orientation of the $\beta^3$-side chain for **5a**, **9b**, and **36**. Dihedral angles between $H_3-C_3-C_4-C_5$ that fit the NMR data best are shown as blue lines, and red lines represent the full MD data set (2500 conformations). RMSD = root-mean-square deviation (for all backbone atoms in structural overlays prepared using the "pair-fit" command in MacPyMOL, Schrödinger Inc.).

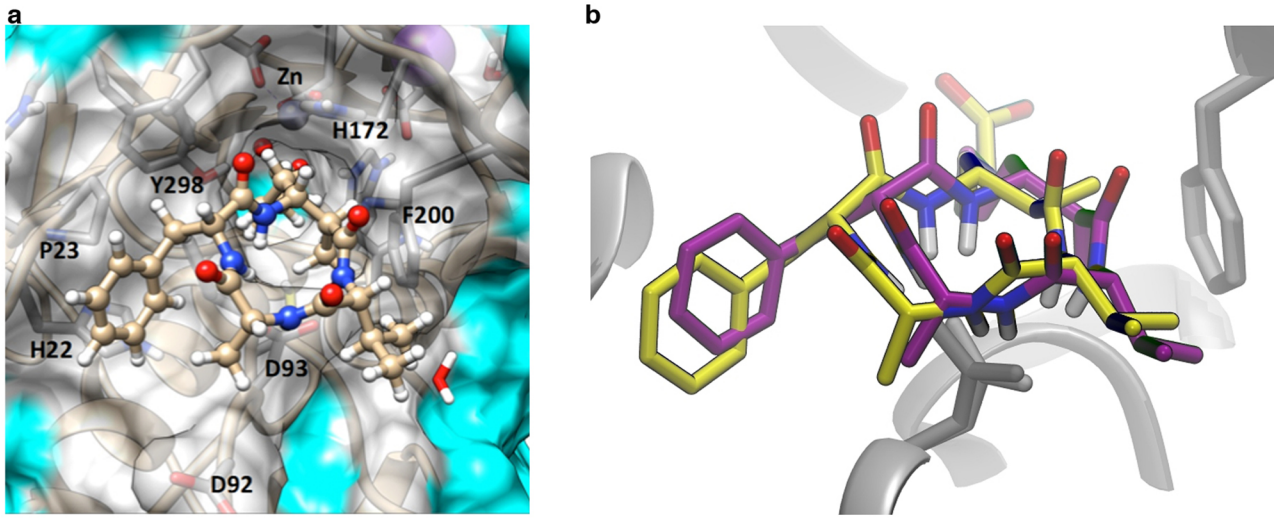
previously synthesized and reported to exhibit less than 50% inhibition across the full panel of HDACs at 50 $\mu$M inhibitor concentrations.[24] Azumamide A (**5a**) was chosen for structure determination over azumamide E (**5e**) due to more well-resolved NMR data, although compound **5a** is less potent on account of its weak carboxamide $Zn^{2+}$-binding group.[20] It can be assumed that this slight structural difference far away from the backbone would not influence the conformation in solution significantly,[38] and thus, compound **5a** should provide a representative three-dimensional structure of the natural products that contain the same macrocycle. Unfortunately, attempts to obtain NMR spectra in $D_2O$ to better simulate the environment in biological assays were hampered by low solubility even after addition of up to 15% (v/v) DMSO-$d_6$; however, subsequent molecular dynamics (MD) simulations in DMSO and water did not show any major differences in conformational space (vide infra). Minor slowly exchanging conformations were indicated in all spectra, presumably due to a *cis*-amide between D-Ala and D-Phe/D-Tyr/D-Trp in agreement with an NOE between $H_{12}$ and $H_{21}$ in azumamide A. The 3D structures of these minor conformations were not determined for any of the compounds due to low signal-to-noise ratios.

Interproton distances were established using the isolated spin-pair approximation (ISPA)[39,40] and *J*-couplings were obtained from proton 1D and DQF-COSY spectra acquired in DMSO-$d_6$ (Supporting Information, Tables S2–S11). The conformation of each analogue that best fit the NMR data was obtained from unrestrained MD simulation in explicit water, and these structures show negligible differences in the peptide backbone (Figure 3b–d) in accordance with comparable *J*-couplings as well as the relative nuclear Overhauser effect (NOE) and rotating-frame nuclear Overhauser effect (ROE). The structure of compound **5a** was essentially the same as the previously reported NMR structure of azumamide E (**5e**) determined in DMSO,[37] except for rotation of the amide between D-Ala and D-Phe (gray and purple structures in Figure 3b). We therefore determined an NOE/ROE-restrained NMR structure (cyan structure in Figure 3b), which resembled the previously reported structure of azumamde E but exhibited inferior overall fit to our NMR data compared to the unrestrained structure (Supporting Information, Tables S12 and S13). We attribute this to conformational averaging that favor short distances and hence amplify the importance of the $NH_{(2)}-H_{21}$ NOE. Overlaying the conformations that fit the NMR data best revealed that only one in 20 of these placed the $NH_{(2)}$ proton in the opposite direction compared to our unrestrained structure (Supporting Information, Figure S5a). This strongly suggests that our NOE-restrained structure does not represent the preferred solution conformation adequately.

In conclusion, on the basis of the high-resolution backbone structures (RMSDs 0.32–0.36 Å), the removal or inversion of the $\beta^2$-methyl group is not likely to cause the observed drop in biochemical activity by inducing conformational changes of the macrocycle. The only major differences observed between the NMR structures were the orientations of the more flexible $\beta^3$-side chain as illustrated by the dihedral angle between $H_3-C_3-C_4-C_5$ in Figure 3e (for ensembles, see Supporting Information, Figure S5a). To investigate whether the preference of this side chain projection in three-dimensional space could explain the large differences in HDACi potencies, we therefore performed computational studies including molecular dynamics calculations on the macrocycles in solution and docking to the recently reported HDAC3:SMRT X-ray crystal structures.[41]

**Molecular Dynamics Simulations of Macrocycles.** Side chain and backbone conformational preferences of azumamides (azumamide E (**5e**), desmethyl azumamide E (**9a**), $\beta^2$-epi-azumamide E (**36**), and dimethyl azumamide E (**8a**)) were addressed by performing molecular dynamics (MD) simulations in both water (TIP3P) and DMSO (Supporting Information, Methods). We measured the dihedral angles


dx.doi.org/10.1021/jm501399d | J. Med. Chem. 2014, 57, 9644–9657




c

| Compound | dG_Bind | | | | | | |
|---|---|---|---|---|---|---|---|
| | Total | Coulomb | Solv. GB | Van der Waals | H-bonding | Lipophilicity | Packing (π–π) |
| AzuE (**5e**) | **–76.2** | –29.3 | 23.7 | –28.6 | –5.2 | –36.5 | –0.3 |
| des-Me-AzuE (**9a**) | **–75.1** | –31.2 | 23.9 | –27.4 | –5.2 | –34.9 | –0.2 |
| epi-AzuE (**36**) | **–71.4** | –29.2 | 24.9 | –27.5 | –5.1 | –34.4 | –0.2 |
| di-Me-AzuE (**8a**) | **–73.9** | –27.4 | 25.5 | –30.4 | –4.8 | –36.6 | –0.1 |

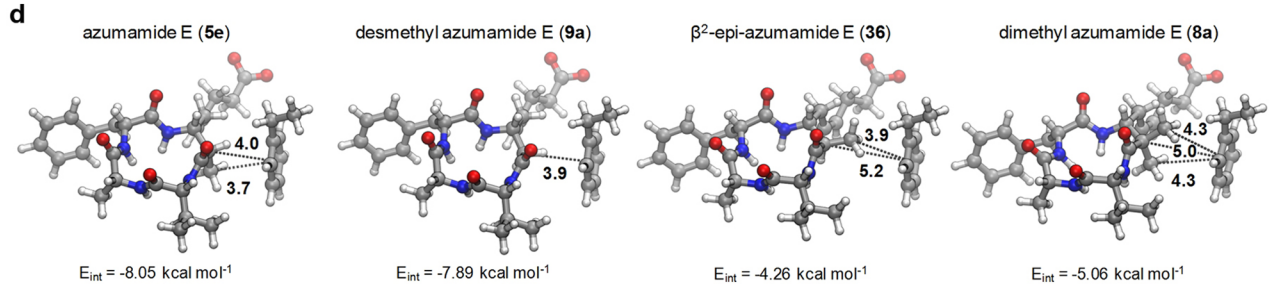



**Figure 4.** Computational results and illustrations. (a) Glide pose for azumamide E docked into the HDAC3B enzyme (PDB ID: 4A69). Azumamides are shown in ball and stick representation, enzyme residues within 6 Å are shown in tube representation beneath a partially transparent gray surface, and the remaining protein surface is in cyan (additional poses including azumamide E analogues are shown in Supporting Information, Figure S13 and S14). (b) Ovelay of docking poses for azumamide E (**5e**, purple) and $\beta^2$-epi-azumamide E (**36**; yellow) shown as sticks. The proteins are shown in gray scale in cartoon format with Asp93 and Phe200 side chains highlighted as sticks. (c) Prime/MMGBSA binding free energies (dG) and energy decomposition for Embrace-optimized geometries for complexes of the azumamide E (AzuE) analogues and HDAC3B. Values are in kcal/mol based on "frozen receptor and frozen ligand" (see Supporting Information, Table S16 for details). (d) Truncated "large" models of docked complexes of HDAC3B and azumamide E and analogues used for DFT calculations of the interaction energy between Phe200 and the azumamide $\beta^2$-position.

H–$C_3$–N–H, H–$C_{12}$–N–H, H–$C_{21}$–N–H, and H–$C_{24}$–N–H, which determine the orientation of the four backbone NH vectors, and histograms shown in Supporting Information, Figure S6–S9 revealed excellent agreement with the NMR data for both water and DMSO simulations. The rotation of the valine side chain ($^i$Pr), given by the dihedral angle $H_{24}$–$C_{24}$–$C_{25}$–$H_{25}$ (Supporting Information, Figures S5b and S10), and the phenylalanine side chain, defined by the $H_{12}$–$C_{12}$–$C_{13}$–$C_{14}$ dihedral angle (Supporting Information, Figure S5c and S11), were also investigated. Finally, projection of the extended $\beta^3$-amino acid side chain, defined by the $H_3$–$C_3$–$C_4$–$C_5$ dihedral angle, was determined from 250 ns MD simulations in water and DMSO for azumamide E (**5e**), desmethyl azumamide E (**9a**), $\beta^2$-epi-azumamide E (**36**), and dimethyl azumamide E (**8a**) as shown in Supporting Information, Figure S12. For azumamide E (**5e**), the major populations were 60 ± 20° (55% of simulation time), −50 ± 20° (33%), and −150 ± 20° (6%) placing the $C_4$–$C_5$ vector projections *trans* to the $\beta^2$-methyl group in the two major populations, and *cis* in the smallest population, which may be explained by steric congestion of the methyl group. This distribution was in agreement with the NMR solution structure of azumamide A (**5a**; Figure 3e), which is identically substituted on the macrocycle (vide supra).

**Docking.** HDAC3 was chosen for docking studies because the available high-resolution X-ray crystal structure of HDAC3 is a complex with its corepressor,[41] and inhibition of HDAC activity has been shown to vary between free and complexed enzymes.[42] Docking into HDAC3A and HDAC3B were performed using Glide (Supporting Information, Methods and Table S14). The docked structure of azumamide E in HDAC3B (Figure 4a and Supporting Information, Figure S13) showed typical features shared by a homogeneous group of poses where the backbone carbonyl oxygen atoms were exposed to solvent and the backbone amide hydrogens were


dx.doi.org/10.1021/jm501399d | J. Med. Chem. 2014, 57, 9644–9657

directed toward the carboxylate side chain of Asp93 in HDAC3. The resulting hydrogen bonds appear to play a key role in favorable positioning of the macrocycles for additional interactions with HDAC3. Most importantly, the $\beta^3$-side chain was located inside the tunnel leading to the HDAC active site and the carboxylate moiety formed a bidentate interaction with the $Zn^{2+}$ ion. The rotation of the $\beta^3$-side chain relative to the azumamide macrocycles, described by the $H_3-C_3-C_4-C_5$ dihedral angles, were either $\sim 180 \pm 20°$ or $\sim 60 \pm 20°$ except for the complex of azumamide E:HDAC3A where it was $\sim 110°$ (Supporting Information, Table S14). This is different from the preferred conformations of this side chain observed by both NMR analysis (Figure 3e) and MD simulations (Supporting Information, Figure S12), suggesting that the inherently preferred projection of this side chain in solution does not affect binding affinity significantly.

The azumamide valine side chain ($^i$Pr) was solvent exposed, and its rotation was unhindered by HDAC residues. The aromatic phenylalanine (tyrosine in azumamide C analogues) side chain was associated with the part of the binding pocket surface comprised of His22 and Pro23. The orientation of the aromatic ring described by the dihedral angle $H_{12}-C_{12}-C_{13}-C_{14}$ varied from $-48°$ to $-97°$, rendering it partially buried between the azumamide core and the HDAC surface. The docked side chain orientations agree with one of the two populations ($H_{12}-C_{12}-C_{13}-C_{14} = -50°$) that were obtained by MD simulation of the unbound azumamide E analogues in explicit water (Supporting Information, Figure S11) and confirmed by NMR analysis (Supporting Information, Figure S5c). The additional orientation indicated by NMR ($\sim 60°$) was prohibited in docked complexes due to collision with Leu266 and Pro23.

The azumamide $\beta^2$-position was located in close proximity to Phe200 in HDAC3, and substitutions at this position therefore affect the interaction with Phe200, giving rise to different poses (Supporting Information, Figure S14). To examine if the observed displacements were artifacts of the rigid-receptor assumption in Glide, we performed postdocking minimization of the Glide poses with Embrace, allowing relaxation of both receptor and ligand. The Embrace-optimized poses again showed substrate dislocations away from the Phe200 side chain for $\beta^2$-epi-azumamide (**36**) and dimethyl azumamide (**8a**) but negligible displacements for azumamide E (**5e**) and desmethyl azumamide E (**9a**) (Figure 4b, Supporting Information, Figure S14 and Table S15). We then proceeded to estimate the binding energies for the azumamide E analogues based on the poses generated with Embrace. As the Glide docking score (XPGscore) is not suitable for quantitative binding free energy estimation, showing errors up to ~4 kcal/mol (http://www.schrodinger.com/kb/793), we estimated binding free energies (dG_Bind) with Prime/MMGBSA calculations on the Embrace-optimized complexes (Supporting Information, Table S16). As shown in Figure 4c, the values of dG_Bind were in semiquantitative agreement with experimental inhibition constants and the decomposition into free energy components showed meaningful trends. For example, the dG_bind "packing" term (for $\pi-\pi$ interactions) quantified the observations from the docked structures to give the ordering azumamide E (**5e**) < des-Me-azumamide E (**9a**) ≈ $\beta^2$-epi-azumamide E (**36**) < di-Me-azumamide E (**8a**), which was in agreement with biochemical potencies.

**Examination of $\beta^2$–Phe200 Interactions by DFT Calculations.** Because of the pivotal role of substitution at the $\beta^2$-position, we then examined the interaction between Phe200 and the $\beta^2$-substitution in azumamide E and analogues using DFT calculations on "small" and "large" model systems obtained by truncation of the Embrace-optimized complexes (see Methods section for details). The DFT energies showed that the Phe200–$\beta^2$-position interaction was always favorable and varied significantly upon substitution at the $\beta^2$-position (Supporting Information, Table S17). Omitting the empirical dispersion correction gave mainly positive interaction energies, confirming the dispersive origin of the attraction as reported for ab initio studies on CH–$\pi$ interactions.[43] The "large" model systems (shown in Figure 4d) revealed Phe200–$\beta^2$-carbon distances for azumamide E (**5e**) and desmethyl azumamide E (**9a**) of 4.0 and 3.9 Å, respectively, which is in agreement with attractive interactions between CH–$\pi$ centers reported for high-resolution crystal structures (peaked at 4.2 Å).[44] Additional stabilization for azumamide E could be explained by placement of the $\beta^2$-methyl group above the center of the Phe200 ring (distance 3.7 Å). The $\beta^2$-carbon atoms of $\beta^2$-epi-azumamide (**36**) and dimethyl azumamide (**8a**), on the other hand, were again substantially displaced from the center of phenyl group (5.2 and 5.0 Å, respectively), preventing dispersive interactions with the $\beta^2$-carbon. Whereas distances of their methyl substituents were within a range allowing for dispersive interactions (3.9 and 4.3 Å), the overall DFT interaction energies were less favorable for these models (Figure 4d).

**Largazole:HDAC8 Cocrystal and other Class I HDAC Crystal Structures.** The docked poses were compared to the crystal structure of the complex between HDAC8 and largazole (PDB ID: 3RQD),[45] which is the only available X-ray crystal structure of a macrocyclic inhibitor in complex with an HDAC enzyme. This revealed a number of similarities, with the macrocycle arranged at the rim of the binding site and the amide hydrogen atoms of largazole facing Asp101 in the protein (analogous to Asp93 in HDAC3). After structural superposition, the Asp101 backbone in HDAC8 coincided with that of Asp93 in HDAC3, although the side chain was differently oriented and formed only one hydrogen bond with a backbone NH in largazole.[45] Thus, a lack of robust interaction between backbone NH groups and Asp101 seems to contribute significantly to the poor inhibitory activities of cyclopeptides against HDAC8, which is supported by the existence of HDAC8 crystal structures that show substantial disorder in the loop region harboring Asp101 (PDB ID: 3SFF).[46] Docking of azumamides into this HDAC8 structure failed while docking into the intact structure (PDB ID: 3RQD)[45] furnished poses where the azumamide ring was arranged approximately as in the HDAC3 dockings, albeit with perturbed backbone amide dihedral angles. Alignment of HDAC1–3 X-ray crystal structures, on the other hand, showed very similar geometries in the region containing this Asp residue (Supporting Information, Figure S15), in agreement with the strong HDAC inhibitory potencies recorded against these isozymes. To briefly address whether the interactions uncovered in this work are generally important, we also docked largazole and apicidin into HDAC3A and B and the obtained poses indeed showed similar geometry and hydrogen bonding interactions with the Asp93 residue as found to be important for azumamide binding (Supporting Information, Methods and Figure S16).

Finally, the natural product romidepsin (**2**) was docked into HDAC3B as shown in Supporting Information, Figure S17 to

test whether the requirements determined for binding of azumamides would apply to this FDA approved drug. As romidepsin lacks a substituent at the site corresponding to position 2 in the azumamides, there is no clash with Phe200 of HDAC3 and thus no favorable dispersive effects either. Nevertheless, the poses for both docked conformations of romidepsin were similar and shared several key features with the favorable poses for azumamides, including protrusion of the $Zn^{2+}$-binding side chain into the channel, leading to the catalytic metal ion, interaction between the thiol and $Zn^{2+}$, as well as peptide backbone hydrogens to the Asp93 side chain of HDAC3.

A more comprehensive series of computational dockings of natural cyclopeptides to X-ray crystal structures of all the class I HDACs is currently underway in our laboratories but is beyond the scope of this report.

**Cytotoxic Activity against Cultured Cancer Cells.** Reversal of chemoresistance caused by silencing of the proapoptotic Bcl-2-family gene, *Bim*, in Burkitt's lymphoma cells has been demonstrated upon treatment with the HDAC inhibitor SAHA.[47] We therefore tested whether a selection of our compounds could influence growth of the Epstein−Barr virus (EBV)-infected human Burkitt's lymphoma cell line, EB-3. Cell viability upon treatment with azumamides B (**5b**), C (**5c**), and E (**5e**), *trans*-olefin (**6b**), saturated compound (**7b**), desmethyl analogues **9a**−**c** as well as SAHA (**1**) and apicidin (**4a**) as controls was determined using a standard MTT assay (Supporting Information, Methods). While SAHA and apicidin exhibited $GI_{50}$ values of 2.8 and 0.8 $\mu$M, respectively, none of the azumamides or analogues exhibited growth inhibition (GI) at concentrations up to 25 $\mu$M, which is surprising considering the potent HDAC inhibitory activities exhibited by the natural products.

## ■ DISCUSSION AND CONCLUSIONS

Naturally occurring macrocyclic peptide-based inhibitors of HDAC enzymes have historically played important roles in the HDAC field. This compound class provides ideal starting points for structure-based investigation of inhibitor−enzyme interactions due to their well-defined three-dimensional structures and inherent modular assembly, which enable ready structural manipulation through chemical synthesis. In this work, we investigated the structural requirements for potent HDAC inhibition by the azumamides and analogues in detail. We first developed efficient syntheses of a series of building blocks, which enabled the assembly of a diverse series of analogues modified to address specific structure−activity relationships. This compound collection was then evaluated by biochemical profiling against the panel of recombinant human HDACs, high-resolution NMR-based determination of selected solution structures, computational evaluation of inhibitor−enzyme complexes, and finally cytotoxic activities against cultured cancer cells in vitro. Through this process we gained valuable insight at a high level of detail into Nature's way of inhibiting this important enzyme class.

The profiling of our compound series revealed only subtle effects in response to the changes made to mimic the saturation/unsaturation of the linker regions of romidepsin, largazole, and apicidin. The identity of aromatic residues mimicking those in apicidin and trapoxin, interestingly, gave rise to opposite effects on potency depending on the enzyme. This is a finding that may be utilized for improving selectivity profiles of macrocyclic HDAC inhibitors, although a maximum difference of 3-fold was observed. By far the most significant effects were found when structurally modifying the 2-position in the $\beta$-amino acid residue. In fact, the changes in activity of 1−2 orders of magnitude resulting from altered methylation at this $\beta^2$-position may be considered a prominent naturally evolved example of the so-called "methyl effect" elegantly discussed by Jorgensen and co-workers for drug-like compounds.[48] We were therefore interested in determining the root of this effect in our system and first hypothesized that the main contribution was conformational remodeling of the azumamide backbone. Somewhat to our surprise, the high-resolution NMR structures of representative ligands revealed no significant conformational variation in the macrocyclic backbone region. In contrast, both NMR *J*-couplings and MD simulations (in explicit water or DMSO) indicated differences in the $\beta^3$-side chain dihedral angles due to steric effects of the $\beta^2$-substituent. Docking to HDAC3 X-ray crystal structures, however, showed that different orientations of the $\beta^3$-side chain enabled coordination to the active site $Zn^{2+}$ ion without compromising key hydrogen bond interactions, which is not entirely surprising considering the degree of flexibility expected for this region of the molecule. Furthermore, the individual docking poses were supported by estimation of binding energies using Prime/MMGBSA. Thus, despite intuitive appeal, a major effect of the $\beta^2$-substitution pattern on favorable preorganization of azumamide conformation appears unlikely. Instead, distorted complexes of epi-methyl-containing analogues (**8a** and **36**) due to steric displacement by Phe200 of HDAC3 resulted in partial disruption of hydrogen bonds to Asp93 as well as overall decrease in favorable dispersive interactions explaining their poor potencies. Because this could not explain the decreased activity of the desmethyl analogues, we performed ab initio calculations on truncated model systems to address the possible effect of a $\beta^2$-methyl−Phe200 interaction in HDAC3, revealing a favorable contribution for the natural product. In summary, our analysis show that requirements for potent inhibition appear to involve (i) $Zn^{2+}$-coordination by protrusion of the $\beta^3$-side chain into the active site tunnel, (ii) optimal hydrogen bonding to Asp93 (Asp101 in HDAC8), and (iii) dispersive interaction with Phe200 in HDAC3. None of the synthetic analogues were able to satisfy all criteria as well as the natural products were, which gave rise to the pronounced methyl effect observed on inhibitory activity. Our detailed analysis thus reveal an intriguing example of a series of compounds where the enthalpic contribution to binding affinity is optimized without suffering prototypical entropic compensation,[49] presumably due to the inherent rigidity of the cyclopeptide backbone.

Moreover, alignment of HDAC1−3 X-ray crystal structures as well as docking of additional macrocyclic HDAC inhibitors strongly indicate that these requirements are broadly important for HDAC inhibition by cyclopeptides. This suggests that drastic structural changes are necessary for these scaffolds to be successfully elaborated into inhibitors exhibiting isozyme selectivity between HDAC1−3, while these features may well be exploited to achieve high HDAC1−3 potency and improved class selectivity.

It is intriguing that Nature has provided these compounds, which seem carefully optimized to bind certain members of the HDAC family of enzymes but do not exhibit any activity when dosed at meaningful concentrations in a cell-based assay. A conceivable explanation for the difference in cellular activity between azumamides C and E compared to SAHA and apicidin, although they exhibit similar in vitro HDACi profiles, could

relate to unfavorable cell-penetrating properties caused by the carboxylate functionality in azumamides. Whether natural prodrug versions of these compounds exist, as in the case of largazole, can only be speculated upon. Nonetheless, ongoing research in our laboratories is aimed at preparation of ester- or thioester-based prodrugs of azumamides to overcome poor intracellular delivery before we further investigate the effects of these chemotypes in cell-based assays.

## EXPERIMENTAL SECTION

Additional synthetic procedures, compound characterization, NMR spectra, and additional methods are provided in the Supporting Information.

**General.** All chemicals and solvents were analytical grade and used without further purification. Vacuum liquid chromatography (VLC) purification was performed on silica gel 60 (particle size 0.015−0.040 μm). UPLC−MS analyses were performed using a Waters Acquity ultra-high-performance liquid chromatography system. A gradient with eluent I (0.1% HCOOH in water) and eluent II (0.1% HCOOH in acetonitrile) rising linearly from 0% to 95% of II during $t$ = 0.00−2.50 min was applied at a flow rate of 1 mL/min (gradient A) or during $t$ = 0.00−5.20 min (gradient B). All compounds were determined by analytical HPLC analysis to be >95% pure on a [150 mm × 4.6 mm, $C_{18}$ Phenomenex Luna column (3 μm)] using an Agilent 1100 LC system equipped with a UV detector. The gradient consisted of eluent III (95:5:0.1, water−MeCN−TFA) and eluent IV (0.1% TFA in acetonitrile) rising linearly from 0% to 95% of IV during $t$ = 2−20 min at a flow rate of 1 mL/min (gradient C) or using a gradient rising linearly from 0% to 40% of IV during $t$ = 5−18 min, then from 40% to 95% of IV during $t$ = 18−20 min at a flow rate of 1 mL/min (gradient D). All final peptides were purified by preparative reversed-phase HPLC on a 250 mm × 20 mm, $C_{18}$ Phenomenex Luna column (5 μm, 100 Å) using an Agilent 1260 LC system equipped with a diode array UV detector and an evaporative light scattering detector (ELSD). A gradient using eluent III and IV rising linearly from 0% to 95% of IV during $t$ = 5−45 min was applied at a flow rate of 20 mL/min (gradient E), and fractions containing pure compound were then pooled and lyophilized. Final compounds were kept at −20 °C. 1D and 2D NMR spectra were recorded on a Varian INOVA 500 MHz instrument, a Bruker Ascend 400 MHz, or a Varian Mercury 300 instrument. All spectra were recorded at 298 K. For the Varian INOVA 500 MHz instrument and the Varian Mercury 300 instrument, the 1D NMR spectra were recorded at 499.9 and 300 MHz for $^1$H and 100 and 75 MHz for $^{13}$C, respectively. The correlation spectroscopy (COSY) spectra were recorded with a relaxation delay of 1.5 s before each scan, a spectral width of 6k × 6k, collecting 8 FIDs and 1k × 512 data points. Heteronuclear single quantum coherence (HSQC) spectra were recorded with a relaxation delay of 1.5 s before each scan, a spectral width of 6k × 25k, collecting 16 FIDs and 1k × 128 data points. Heteronuclear 2-bond correlation (H2BC) spectra were recorded with a relaxation delay of 1.5 s before each scan, a spectral width of 4k × 35k, collecting 16 FIDs at 295 K and 1k × 256 data points. Heteronuclear multiple-bond correlation (HMBC) spectra were recorded with a relaxation delay of 1.5 s before each scan, a spectral width of 6k × 35k, collecting 32 FIDs and 1k × 256 data points. Finally, on the Bruker Ascend 400 MHz, the 1D NMR spectra were recorded at 400 MHz for $^1$H and 100 MHz for $^{13}$C. The correlation spectroscopy (COSY) spectra were recorded with a relaxation delay of 1.5 s before each scan, a spectral width of 3k × 3k, collecting 4 FIDs and 1k × 128 data points. The heteronuclear single quantum coherence (HSQC) spectra were recorded with a relaxation delay of 1.5 s before each scan, a spectral width of 4800 × 16600, collecting 4 FIDs and 1k × 256 data points. Chemical shifts are reported in ppm relative to deuterated solvent peaks as internal standards (δH, DMSO-$d_6$ 2.50 ppm; δC, DMSO-$d_6$ 39.52 ppm; δH, $CD_3OD$ 3.31 ppm; δH, $CDCl_3$ 7.26 ppm; δC, $CDCl_3$ 77.16 ppm). Coupling constants ($J$) are given in hertz (Hz). Multiplicities of $^1$H NMR signals are reported as follows: s, singlet; br s, broad singlet; d, doublet; t, triplet; q, quartet; m, multiplet.

*E)-6-((2R,5R,8R,11S)-8-Benzyl-2-isopropyl-5-methyl-3,6,9,13-tetraoxo-1,4,7,10-tetraazacyclotridecan-11-yl)hex-5-enoic Acid (**6a**).* Compound **13** (31.3 mg, 0.1 mmol), HATU (38.8 mg, 0.1 mmol, 1 equiv), and 2,6-lutidine (23 μL, 0.2 mmol, 2 equiv) in DMF (1.5 mL) were preincubated for 5 min and added to a tripeptide (D-Val-D-Ala-D-Phe) on solid support. After 18 h, the peptide was washed with DMF (×3), MeOH (×3), and $CH_2Cl_2$ (×3). The peptide was cleaved with 50% TFA in $CH_2Cl_2$ (2 × 30 min) and concentrated in vacuo. Trituration in $Et_2O$ (8 mL) gave a crude tetrapeptide (88 mg), which was dissolved in DMF (130 mL). Addition of $i$-$Pr_2NEt$ (140 μL, 0.8 mmol, 8 equiv) was followed by HATU (78 mg, 0.2 mmol, 2 equiv), and the reaction was stirred for 66 h and then concentrated in vacuo. LiOH (32 mg, 1.3 mmol, 13 equiv) in THF−$H_2O$ (4 mL, 1:1) was added to the crude material, and the mixture was stirred for 7 h. Preparative HPLC gave the desired product (3.8 mg, 7%, over 4 steps) as a white solid. Two conformations were observed in 89:11 ratio determined by integration of $^1$H NMR peaks. Characterization is given for the major conformation. $^1$H NMR (500 MHz, DMSO-$d_6$) δ 12.03 (s, 1H), 7.87 (d, $J$ = 6.8 Hz, 1H), 7.61 (d, $J$ = 9.7 Hz, 1H), 7.40 (d, $J$ = 7.5 Hz, 2H), 7.33−7.11 (m, 5H), 5.46−5.41 (m, 2H), 4.41−4.31 (m, 1H), 4.31−4.24 (m, 2H), 3.82 (t, $J$ = 6.6 Hz, 1H), 3.06−2.81 (m, 2H), 2.62−2.48 (m, 2H), 2.20 (t, $J$ = 7.4 Hz, 2H), 2.07−1.97 (m, 2H), 1.97−1.87 (m, 1H), 1.66−1.48 (m, 2H), 1.15 (d, $J$ = 7.3 Hz, 3H), 0.92 (dd, $J$ = 9.0, 7.0 Hz, 6H). HRMS calcd for $C_{26}H_{36}N_4O_6H^+$ $[M + H]^+$, 501.2708; found, 501.2716. HPLC gradient C, $t_R$ = 11.6 min (HPLC purity >99%, 230 nm).

*(E)-6-((2R,5R,8R,11S)-8-(4-Hydroxybenzyl)-2-isopropyl-5-methyl-3,6,9,13-tetraoxo-1,4,7,10-tetraazacyclo-tridecan-11-yl)hex-5-enoic Acid (**6b**).* Compound **13** (31.3 mg, 0.1 mmol), HATU (38.8 mg, 0.1 mmol, 1 equiv) and 2,6-lutidine (23 μL, 0.2 mmol, 2 equiv) in DMF (1.5 mL) were preincubated for 5 min and added to a tripeptide (D-Val-D-Ala-D-Tyr) on solid support. After 18 h, the peptide was washed with DMF (×3), MeOH (×3) and $CH_2Cl_2$ (×3). The peptide was cleaved with 50% TFA in $CH_2Cl_2$ (2 × 30 min) and concentrated in vacuo. Trituration in $Et_2O$ (8 mL) gave a crude tetrapeptide (89 mg), which was dissolved in DMF (130 mL). Addition of $i$-$Pr_2NEt$ (140 μL, 0.8 mmol, 8 equiv) was followed by HATU (78 mg, 0.2 mmol, 2 equiv). The reaction was stirred for 66 h and then concentrated in vacuo. An unsuccessful purification provided a crude methylester (7.2 mg), which was dissolved in THF−$H_2O$ (4 mL, 1:1). LiOH (1.3 mg, 0.054 mmol, ∼4 equiv) was added, and the mixture was stirred for 18 h. Additional LiOH (4 mg, 0.15 mmol, ∼8 equiv) was added, and the mixture was stirred for 7 h at room temperature, then 63 h at 5 °C. Preparative HPLC gave the desired product (4.2 mg, 8% over 4 steps) as a white solid. Two conformations were observed in 89:11 ratio determined by integration of $^1$H NMR peaks. Characterization is given for the major conformation. $^1$H NMR (500 MHz, DMSO-$d_6$) δ 9.23 (s, 1H), 7.85 (d, $J$ = 6.7 Hz, 1H), 7.52 (d, $J$ = 9.7 Hz, 1H), 7.38 (d, $J$ = 9.9 Hz, 1H), 7.31 (d, $J$ = 7.3 Hz, 1H), 6.95 (d, $J$ = 8.4 Hz, 2H), 6.66 (d, $J$ = 8.4 Hz, 2H), 5.47−5.39 (m, 2H), 4.40−4.33 (m, 1H), 4.29 (dq, $J$ = 14.9, 7.6 Hz, 1H), 4.19 (td, $J$ = 9.3, 6.0 Hz, 1H), 3.81 (t, $J$ = 6.5 Hz, 1H), 2.84 (dd, $J$ = 13.9, 5.8 Hz, 1H), 2.75 (dd, $J$ = 13.8, 9.1 Hz, 1H), 2.58−2.40 (m, 2H), 2.20 (t, $J$ = 7.4 Hz, 2H), 2.11−1.98 (m, 2H), 1.98−1.90 (m, 1H), 1.65−1.48 (m, 2H), 1.16 (d, $J$ = 7.4 Hz, 3H), 0.92 (dd, $J$ = 8.3, 7.2 Hz, 6H). HRMS calcd for $C_{26}H_{36}N_4O_7H^+$ $[M + H]^+$, 517.2657; found, 517.2660. HPLC gradient C, $t_R$ = 10.3 min (HPLC purity >99%, 230 nm).

*(E)-6-((2R,5R,8R,11S)-8-((1H-Indol-2-yl)methyl)-2-isopropyl-5-methyl-3,6,9,13-tetraoxo-1,4,7,10-tetraazacyclotridecan-11-yl)hex-5-enoic Acid (**6c**).* Compound **13** (31.3 mg, 0.1 mmol), HATU (38.8 mg, 0.1 mmol, 1 equiv), and 2,6-lutidine (23 μL, 0.2 mmol, 2 equiv) in DMF (1.5 mL) were preincubated for 5 min and added to a tripeptide (D-Val-D-Ala-D-Trp) on solid support. After 18 h, the peptide was washed with DMF (×3), MeOH (×3), and $CH_2Cl_2$ (×3). The peptide was cleaved with 50% TFA in $CH_2Cl_2$ (2 × 30 min) and concentrated in vacuo. Trituration in $Et_2O$ (8 mL) gave a crude tetrapeptide, which was dissolved in DMF (130 mL). Addition of $i$-$Pr_2NEt$ (140 μL, 0.8 mmol, 8 equiv) was followed by HATU (78 mg, 0.2 mmol, 2 equiv). The reaction was stirred for 66 h and then concentrated in vacuo. An unsuccessful purification provided a crude methylester (2.9 mg), which


dx.doi.org/10.1021/jm501399d | J. Med. Chem. 2014, 57, 9644−9657

was dissolved in THF−$H_2O$ (4 mL, 1:1). LiOH (0.5 mg, 0.02 mmol, ∼4 equiv) was added, and the mixture was stirred for 18 h. Additional LiOH (1 mg, 0.04 mmol, ∼8 equiv) was added, and the mixture was stirred for 6 h. Preparative HPLC gave the desired product (1.4 mg, 3% over four steps) as a white solid. Two conformations were observed in 92:8 ratio determined by integration of $^1$H NMR peaks. Characterization is given for the major conformation. $^1$H NMR (500 MHz, DMSO-$d_6$) $\delta$ 7.82 (d, $J$ = 6.8 Hz, 1H), 7.55 (d, $J$ = 9.7 Hz, 1H), 7.50 (d, $J$ = 7.8 Hz, 1H), 7.45 (d, $J$ = 9.8 Hz, 1H), 7.37 (d, $J$ = 7.3 Hz, 1H), 7.34 (d, $J$ = 8.1 Hz, 1H), 7.16−7.04 (m, 2H), 6.98 (t, $J$ = 7.5 Hz, 1H), 5.47−5.31 (m, 2H), 4.45−4.24 (m, 3H), 3.82 (t, $J$ = 6.5 Hz, 1H), 3.09−3.00 (m, 2H), 2.48−2.37 (m, 2H), 2.19 (t, $J$ = 7.3 Hz, 2H), 2.06−1.87 (m, 4H), 1.59−1.46 (m, 2H), 1.17 (d, $J$ = 7.3 Hz, 3H), 0.92 (t, $J$ = 6.4 Hz, 6H). HRMS calcd for $C_{28}H_{37}N_5O_6H^+$ [M + H]$^+$, 540.2817; found, 540.2807. HPLC gradient C, $t_R$ = 11.7 min (HPLC purity >99%, 230 nm).

*6-((2R,5R,8R,11R)-8-Benzyl-2-isopropyl-5-methyl-3,6,9,13-tetraoxo-1,4,7,10-tetraazacyclotridecan-11-yl)hexanoic Acid (**7a**).* Compound **14** (25.4 mg, 0.08 mmol), HATU (36.5 mg, 0.1 mmol, 1.3 equiv), and 2,6-lutidine (18 $\mu$L, 0.16 mmol, 2 equiv) in DMF (1.5 mL) were preincubated for 5 min and added to a tripeptide (D-Val-D-Ala-D-Phe) on solid support. After 18 h, the peptide was washed with DMF (×3), MeOH (×3), and $CH_2Cl_2$ (×3). The peptide was cleaved with 50% TFA in $CH_2Cl_2$ (2 × 30 min.) and concentrated in vacuo. Trituation in $Et_2O$ (8 mL) gave a crude tetrapeptide, which was dissolved in DMF (170 mL). Addition of $i$-$Pr_2NEt$ (140 $\mu$L, 0.8 mmol, 10 equiv) was followed by HATU (61 mg, 0.16 mmol, 2 equiv). The reaction was stirred for 24 h and then concentrated in vacuo. LiOH (34 mg, 1.5 mmol, ∼20 equiv) in THF−$H_2O$ (15 mL, 1:2) was added to the crude compound and the mixture stirred for 4 h. Additional LiOH (19 mg, 0.8 mmol, ∼10 equiv) was added, and the reaction was stirred for 6 h. Then, LiOH (35 mg, 1.5 mmol, ∼20 equiv) was added, and the reaction was stirred for additional 17 h. Preparative HPLC gave the desired product (4.2 mg, 10% over 4 steps) as a white solid. Two conformations were observed in 91:9 ratio determined by integration of $^1$H NMR peaks. Characterization is given for the major conformation. $^1$H NMR (500 MHz, DMSO-$d_6$) $\delta$ 11.97 (s, 1H), 7.78 (d, $J$ = 6.6 Hz, 1H), 7.55 (d, $J$ = 8.7 Hz, 1H), 7.41 (d, $J$ = 8.4 Hz, 1H), 7.34−7.10 (m, 6H), 4.31−4.20 (m, 2H), 3.81 (t, $J$ = 7.0 Hz, 2H), 2.93 (dd, $J$ = 13.7, 6.7 Hz, 1H), 2.84 (dd, $J$ = 13.7, 8.5 Hz, 1H), 2.59−2.45 (m, 1H), 2.41−2.27 (m, 1H), 2.19 (t, $J$ = 7.4 Hz, 2H), 1.94 (dq, $J$ = 13.4, 6.7 Hz, 1H), 1.58−1.21 (m, 8H), 1.17 (d, $J$ = 7.3 Hz, 3H), 0.92 (dd, $J$ = 9.1, 7.0 Hz, 6H). HRMS calcd for $C_{26}H_{38}N_4O_6H^+$ [M + H]$^+$, 503.2864; found, 503.2870. HPLC gradient D, $t_R$ = 18.3 min (HPLC purity >99%, 230 nm).

*6-((2R,5R,8R,11R)-8-(4-Hydroxybenzyl)-2-isopropyl-5-methyl-3,6,9,13-tetraoxo-1,4,7,10-tetraazacyclo-tridecan-11-yl)hexanoic Acid (**7b**).* Compound **14** (25.4 mg, 0.08 mmol), HATU (36.5 mg, 0.1 mmol, 1 equiv), and 2,6-lutidine (18 $\mu$L, 0.16 mmol, 2 equiv) in DMF (1.5 mL) were preincubated for 5 min and then added to a tripeptide (D-Val-D-Ala-D-Tyr) on solid support. After 18 h, the peptide was washed with DMF (×3), MeOH (×3) and $CH_2Cl_2$ (×3). The peptide was cleaved with 50% TFA in $CH_2Cl_2$ (2 × 30 min) and concentrated in vacuo. Trituration in $Et_2O$ (8 mL) gave a crude tetrapeptide, which was dissolved in DMF (170 mL) and treated with $i$-$Pr_2NEt$ (140 $\mu$L, 0.8 mmol, 10 equiv) and HATU (61 mg, 0.16 mmol, 2 equiv). The reaction was stirred for 24 h and then concentrated in vacuo. LiOH (43 mg, 1.8 mmol, 23 equiv) in THF−$H_2O$ (15 mL, 1:2) was added to the crude compound and stirred for 24 h, an additional amount of LiOH (38 mg, 1.6 mmol, 20 equiv) was added, and the reaction was stirred for 24 h. Preparative HPLC gave the desired product (3.1 mg, 7% over 4 steps) as a white solid. Two conformations were observed in 87:13 ratio determined by integration of $^1$H NMR peaks. Characterization is given for the major conformation. $^1$H NMR (500 MHz, DMSO-$d_6$) $\delta$ 9.21 (s, 1H), 7.77 (d, $J$ = 7.0 Hz, 1H), 7.44 (d, $J$ = 9.5 Hz, 1H), 7.37 (d, $J$ = 9.8 Hz, 1H), 7.17 (d, $J$ = 7.4 Hz, 1H), 6.94 (d, $J$ = 8.4 Hz, 2H), 6.65 (d, $J$ = 8.4 Hz, 2H), 4.26 (dq, $J$ = 14.8, 7.3 Hz, 1H), 4.19−4.14 (m, 1H), 4.00−3.74 (m, 2H), 2.80 (dd, $J$ = 13.8, 6.6 Hz, 1H), 2.72 (dd, $J$ = 13.8, 8.4 Hz, 1H), 2.49 (dd, $J$ = 8.5, 6.8 Hz, 1H), 2.32 (dd, $J$ = 14.5, 3.2 Hz, 1H), 2.20 (t, $J$ = 7.3 Hz, 2H), 1.99−1.90 (m, 1H), 1.67−1.21 (m, 8H), 1.18 (d, $J$ = 7.4 Hz, 3H), 0.92 (dd, $J$ = 8.5, 7.1 Hz, 6H). HRMS calcd for $C_{26}H_{38}N_4O_7H^+$ [M + H]$^+$, 519.2813; found, 519.2813. HPLC gradient C, $t_R$ = 10.4 min (HPLC purity >99%, 230 nm).

*6-((2R,5R,8R,11R)-8-((1H-Indol-2-yl)methyl)-2-isopropyl-5-methyl-3,6,9,13-tetraoxo-1,4,7,10-tetraazacyclotridecan-11-yl)hexanoic Acid (**7c**).* The saturated Fmoc-$\beta$-amino acid (see, Supporting Information) (95 mg, 0.22 mmol) was dissolved in anhydrous $CH_2Cl_2$ (5 mL) and $i$-$Pr_2NEt$ (75 $\mu$L, 0.43 mmol, 2 equiv) and loaded onto a 2-chlorotrityl resin. The resin was washed with $CH_2Cl_2$ (×3) and then capped with $i$-$Pr_2NEt$−MeOH−$CH_2Cl_2$ (1:2:7) followed by washing. Theoretical loading of the resin was calculated to be 0.7 mmol/g. Standard Fmoc SPPS using HATU as the coupling reagent was applied to couple D-Trp(Boc), D-Ala, and D-Val. The peptide was cleaved by treatment with AcOH−$CF_3CH_2OH$−$CH_2Cl_2$ (1:1:3) for 2.5 h, and the mixture was then concentrated in vacuo. Trituration in $Et_2O$ (8 mL) gave a crude tetrapeptide, half of which was cyclized in DMF (170 mL) using $i$-$Pr_2NEt$ (100 $\mu$L, 0.6 mmol, ∼6 equiv) and HATU (57 mg, 0.15 mmol, 1.5 equiv). The reaction was stirred for 2 h and then concentrated in vacuo. LiOH (7 mg, 0.3 mmol, 3 equiv) in THF−$H_2O$ (5 mL, 1:1) was added to the crude compound and stirred 5 h, whereafter additional LiOH (48 mg, 2 mmol, 20 equiv) was added and the reaction was stirred for 24 h. The crude mixture was concentrated in vacuo and dissolved in $CH_2Cl_2$ (6 mL) followed by addition of TFA (6 mL). The reaction was stirred for 30 min and then concentrated in vacuo. Preparative HPLC gave the desired product (12.4 mg, 48% over 12 steps) as a white solid. Two conformations were observed in 91:9 ratio determined by integration of $^1$H NMR peaks. Characterization is given for the major conformation. $^1$H NMR (500 MHz, DMSO-$d_6$) $\delta$ 11.97 (s, 1H), 10.81 (d, $J$ = 56.2 Hz, 1H), 7.75 (d, $J$ = 6.8 Hz, 1H), 7.57−7.41 (m, 3H), 7.33 (d, $J$ = 8.0 Hz, 1H), 7.22 (d, $J$ = 7.2 Hz, 1H), 7.10−7.01 (m, 2H), 7.00−6.93 (m, 1H), 4.43−4.32 (m, 1H), 4.28 (dq, $J$ = 15.1, 7.4 Hz, 1H), 3.87−3.77 (m, 2H), 3.04 (dd, $J$ = 14.7, 6.8 Hz, 1H), 2.99 (dd, $J$ = 14.8, 7.6 Hz, 1H), 2.55−2.45 (m, 1H), 2.31 (dd, $J$ = 14.4, 3.3 Hz, 1H), 2.19 (t, $J$ = 7.3 Hz, 2H), 2.03−1.85 (m, 1H), 1.54−1.10 (m, 8H), 1.18 (d, $J$ = 7.4 Hz, 3H), 0.91 (t, $J$ = 6.5 Hz, 6H). HRMS calcd for $C_{28}H_{39}N_5O_6H^+$ [M + H]$^+$, 542.2973; found, 542.2978. HPLC gradient C, $t_R$ = 11.98 min (HPLC purity >99%, 230 nm).

*(Z)-6-((2R,5R,8R,11R)-8-Benzyl-2-isopropyl-5,12,12-trimethyl-3,6,9,13-tetraoxo-1,4,7,10-tetraazacyclotridecan-11-yl)hex-4-enoic Acid (**8a**).* The linear peptide **33a** (35 mg, 0.08 mmol) was dissolved in DMF (200 mL), whereafter $i$-$Pr_2NEt$ (140 $\mu$L, 0.80 mmol, 10 equiv) and HATU (61 mg, 0.16 mmol, 2.0 equiv) were added. The solution was stirred for 16 h, and the reaction mixture was concentrated in vacuo. Addition of EtOAc (200 mL) afforded a white precipitate (HOAt), which was filtered off. The EtOAc was washed with aq HCl (1.0 M, 10 mL × 3) and concentrated in vacuo to afford the crude cyclic peptide (76 mg) as brown oil. UPLC-MS gradient B, $t_R$ = 1.92 min. MS calcd for $C_{32}H_{51}N_4O_6H^+$, 557.7; found, 557.2 [M + H]$^+$. Then, LiOH (67 mg, 2.79 mmol, 35 equiv) in water (5 mL) was added to the crude cyclic peptide in THF (5 mL), and after 16 h, the mixture was concentrated in vacuo. The residue was dissolved in DMF and purified by preparative HPLC to afford the acid **8a** (2.6 mg, 4%, 5 steps) as a white solid. Two conformations were observed in a 91:9 ratio determined by integration of $^1$H NMR peaks. Characterization is given for the major conformation. $^1$H NMR (500 MHz, DMSO-$d_6$) $\delta$ 8.04 (br s, 1H), 7.38 (br s, 1H), 7.27−7.11 (m, 6H), 7.05 (d, $J$ = 9.5 Hz, 1H), 5.16 (m, 1H), 4.99 (m, 1H), 4.40 (m, 1H), 4.14 (dq, $J$ = 8.4, 7.9 Hz, 1H), 4.05 (m, 1H), 3.70 (dd, $J$ = 10.1, 8.7 Hz, 1H), 3.00 (dd, $J$ = 13.6, 8.0 Hz, 2H), 2.70 (dd, $J$ = 13.6, 6.6 Hz, 1H), 2.16 (m, 5H), 2.01 (m, 2H), 1.15−1.3 (m, 6H), 1.04 (s, 3H), 0.85−0.84 (m, 6H). HRMS calcd for $C_{28}H_{40}N_4O_6Na^+$ [M + Na]$^+$, 551.2840; found, 551.2839. HPLC gradient C, $t_R$ = 12.9 min (HPLC purity >95%, 230 nm).

*(Z)-6-((2R,5R,8R,11R)-8-(4-Hydroxybenzyl)-2-isopropyl-5,12,12-trimethyl-3,6,9,13-tetraoxo-1,4,7,10-tetraazacyclotridecan-11-yl)-hex-4-enoic Acid (**8b**).* The linear peptide **33b** (35 mg, 0.092 mmol) was dissolved in DMF (230 mL), whereafter $i$-$Pr_2NEt$ (159 $\mu$L, 0.92 mmol, 10 equiv) and HATU (70 mg, 0.19 mmol, 2.0 equiv) were


dx.doi.org/10.1021/jm501399d | J. Med. Chem. 2014, 57, 9644−9657

added. The solution was stirred for 16 h, and the reaction mixture was concentrated in vacuo. Addition of EtOAc (200 mL) afforded a white precipitate (HOAt), which was filtered off. The EtOAc was washed with aq HCl (1.0 M, 10 mL × 3) and concentrated in vacuo to afford the crude cyclic peptide (123 mg) as brown oil. UPLC-MS gradient B, $t_R$ = 1.72 min. MS calcd for $C_{30}H_{43}N_4O_7^-$ [M − H]$^-$, 571.3; found, 571.3. LiOH (77 mg, 3.2 mmol, 35 equiv) in water (5 mL) was added to the crude cyclic peptide (123 mg, ∼0.08 mmol) in THF (5 mL). After 20 h, the mixture was concentrated in vacuo and the residue was dissolved in DMF and purified by preparative HPLC to afford the acid **8b** (1.0 mg, 2%, 5 steps) as a white solid. Two conformations were observed in 89:11 ratio determined by integration of $^1$H NMR peaks. Characterization is given for the major conformation. $^1$H NMR (500 MHz, DMSO-$d_6$) $\delta$ 8.03 (d, $J$ = 8.6 Hz, 1H), 7.29 (d, $J$ = 8.6 Hz, 1H), 7.19 (d, $J$ = 8.3 Hz, 1H), 7.00 (d, $J$ = 9.9 Hz, 1H), 6.95 (d, $J$ = 8.4 Hz, 2H), 6.60 (d, $J$ = 8.4 Hz, 2H), 5.19 (m, 1H), 5.01 (m, 1H), 4.32 (q, $J$ = 8.0 Hz, 1H), 4.13 (pentet, $J$ = 7.0 Hz, 1H), 4.04 (m, 1H), 3.69 (t, $J$ = 10.1, 8.3 Hz, 1H), 2.87 (dd, $J$ = 13.7, 8.0 Hz, 1H), 2.57 (dd, $J$ = 13.7, 6.6 Hz, 1H), 2.18 (m, 5H), 2.00 (m, 2H), 1.15 (m, 6H), 1.03 (s, 3H), 0.84 (m, 6H). HRMS calcd for $C_{28}H_{40}N_4O_7H^+$ [M + H]$^+$, 545.2970; found, 545.2968. HPLC gradient C, $t_R$ = 11.4 min (HPLC purity >95%, 230 nm).

*(Z)-6-((2R,5R,8R,11R)-8-((1H-Indol-3-yl)methyl)-2-isopropyl-5,12,12-trimethyl-3,6,9,13-tetraoxo-1,4,7,10-tetraazacyclotridecan-11-yl)hex-4-enoic Acid (**8c**).* The linear peptide **33c** (43 mg, 0.059 mmol) was dissolved in DMF (150 mL), whereafter $i$-Pr$_2$NEt (51.4 μL, 0.30 mmol, 5 equiv) and HATU (34 mg, 0.089 mmol, 1.5 equiv) were added. The solution was stirred for 17 h, and the reaction mixture was concentrated in vacuo to afford the crude cyclic peptide. UPLC-MS gradient B, $t_R$ = 1.72 min. MS calcd for $C_{32}H_{45}N_5O_6H^+$ [M + H]$^+$, 595.3; found, 596.3. Then, LiOH (26 mg, 1.09 mmol, 50 equiv) in water (4 mL) was added to the crude cyclic peptide (13 mg, ∼0.022 mmol) in THF (4 mL). After 18 h, the mixture was concentrated in vacuo and the residue was dissolved in MeCN−water−DMF (2:2:1, 2.5 mL) and purified by preparative HPLC to afford the acid **8c** (2.6 mg, 6%, 5 steps) as a white solid. Two conformations are observed in a 92:8 ratio determined by integration of $^1$H NMR peaks. Characterization data are given for the major conformation. $^1$H NMR (500 MHz, DMSO-$d_6$) $\delta$ 7.74 (d, $J$ = 8.6 Hz, 1H), 7.53 (d, $J$ = 7.8 Hz, 1H), 7.39 (d, $J$ = 8.3 Hz, 1H), 7.32 (d, $J$ = 8.1 Hz, 1H), 7.24 (d, $J$ = 8.4 Hz, 1H), 7.16 (d, $J$ = 9.2 Hz, 1H), 7.05 (m, 2H), 6.97 (m, 1H), 5.37−5.24 (m, 2H), 4.14 (m, 2H), 3.95 (td, $J$ = 9.5, 4.1 Hz, 1H), 3.71 (dd, $J$ = 10.0, 8.5 Hz, 1H), 3.08 (d, $J$ = 7.6 Hz, 2H), 2.36−2.17 (m, 6H), 2.08 (m, 1H), 1.32 (s, 3H), 1.23 (d, $J$ = 7.3 Hz, 3H), 1.09 (s, 3H), 0.88 (d, $J$ = 6.9 Hz, 3H), 0.87 (d, $J$ = 7.0 Hz, 3H). HRMS calcd for $C_{30}H_{41}N_5O_6H^+$ [M + H]$^+$, 568.3130; found, 568.3135. HPLC gradient C, $t_R$ = 11.6 min (HPLC purity >95%, 230 nm).

*(Z)-6-((2R,5R,8R,11R)-8-Benzyl-2-isopropyl-5-methyl-3,6,9,13-tetraoxo-1,4,7,10-tetraazacyclotridecan-11-yl)hex-4-enoic Acid (**9a**).* The linear peptide **34a** (45 mg, 0.068 mmol) was dissolved in DMF (200 mL), and $i$-Pr$_2$NEt (131 μL, 0.75 mmol, 10 equiv) and HATU (57 mg, 0.15 mmol, 2.2 equiv) were added. The solution was stirred for 16 h, and the reaction mixture was concentrated in vacuo. Addition of EtOAc (200 mL) afforded a white precipitate (HOAt), which was filtered off. The EtOAc was washed with aqueous HCl (1.0 M, 10 mL × 3) and concentrated in vacuo to afford the crude cyclic peptide (127 mg) as brown oil. UPLC-MS gradient A $t_R$ = 1.87 min. MS calcd for $C_{28}H_{40}N_4O_6H^+$ [M + H]$^+$, 529.3; found, 529.4. Then, LiOH (38 mg, 1.59 mmol) in water (5 mL) was added to the cyclic peptide (127 mg, ∼0.068 mmol) in THF (5 mL). To ensure a pH >10, additional LiOH (21 mg, 0.88 mmol, 13.0 equiv) was added and the solution was stirred for 16 h. The mixture was concentrated in vacuo and dissolved in DMF−MeCN (8:2, 3.0 mL) before purification by preparative HPLC afforded **9a** (8.1 mg, 21%, 5 steps) as white solid. Two conformations are observed in a 94:6 ratio determined by integration of $^1$H NMR peaks. Characterization is given for the major conformation. $^1$H NMR (800 MHz, DMSO-$d_6$) $\delta$ 12.08 (s, 1H), 7.86 (d, $J$ = 6.8 Hz, 1H), 7.66 (d, $J$ = 9.0 Hz, 1H), 7.43 (d, $J$ = 9.0 Hz, 1H), 7.35 (d, $J$ = 6.6 Hz, 1H), 7.28 (m, 2H), 7.16−7.23 (m, 3H), 5.48 (m, 1H), 5.37 (dt, $J$ = 10.5, 7.4 Hz, 1H), 4.30−4.21 (m, 2H), 3.93 (dd, $J$ = 7.2, 3.6 Hz, 1H), 3.81 (t, $J$ = 6.7 Hz, 1H), 2.95 (dd, $J$ = 13.8, 5.8 Hz, 1H), 2.82 (dd, $J$ = 13.8, 9.5 Hz, 1H), 2.50 (m, 1H), 2.29 (m, 7H), 1.94 (m, 1H), 1.15 (d, $J$ = 7.3 Hz, 3H), 0.94 (d, $J$ = 6.8 Hz, 3H), 0.91 (d, $J$ = 6.8 Hz, 3H). HRMS calcd for $C_{26}H_{36}N_4O_6H^+$ [M + H]$^+$, 501.2708; found, 501.2713. HPLC gradient C, $t_R$ = 11.7 min (HPLC purity >95%, 230 nm).

*(Z)-6-((2R,5R,8R,11R)-8-(4-Hydroxybenzyl)-2-isopropyl-5-methyl-3,6,9,13-tetraoxo-1,4,7,10-tetraazacyclotridecan-11-yl)hex-4-enoic Acid (**9b**).* The linear peptide **34b** (67 mg, 0.099 mmol) was dissolved in DMF (240 mL), and $i$-Pr$_2$NEt (138 μL, 0.79 mmol, 8 equiv) and HATU (57 mg, 0.15 mmol, 1.5 equiv) were added. The solution was stirred for 16 h, and the reaction mixture was concentrated in vacuo. Purification by VLC afforded the protected cyclic peptide (22 mg) as a white amorphous solid, which was used without further purification. UPLC-MS gradient B, $t_R$ = 1.60 min. MS calcd for $C_{28}H_{42}N_4O_8H^+$ [M + H]$^+$, 545.3; found, 545.3. LiOH (44 mg, 1.84 mmol, 50 equiv) in water (4 mL) was added to the cyclic peptide (20 mg, 0.037 mmol) in THF (4 mL). After 17 h, the mixture was concentrated in vacuo, dissolved in DMF−MeCN−water (2:1:2, 2.5 mL), and purified by preparative HPLC to afford **9b** (11.2 mg, 22%, 5 steps) as a white solid. Two conformations were observed in a 92:8 ratio determined by integration of $^1$H NMR peaks. Characterization is given for the major conformation. $^1$H NMR (500 MHz, DMSO-$d_6$) $\delta$ 12.07 (br s, 1H), 7.87 (d, $J$ = 6.9 Hz, 1H), 7.56 (d, $J$ = 9.7 Hz, 1H), 7.43 (d, $J$ = 9.8 Hz, 1H), 7.25 (d, $J$ = 7.0 Hz, 1H), 6.95 (d, $J$ = 8.4 Hz, 2H), 6.65 (d, $J$ = 8.4 Hz, 2H), 6.52 (br s, 1H), 5.47 (m, 1H), 5.37 (m, 1H), 4.28 (dq, $J$ = 9.7, 7.4 Hz, 1H), 4.16 (m, 1H), 3.93 (m, 1H), 3.81 (t, $J$ = 6.6 Hz, 1H), 2.82 (dd, $J$ = 13.9, 5.7 Hz, 1H), 2.70 (dd, $J$ = 13.9, 9.3 Hz, 1H), 2.47 (m, 1H), 2.27 (m, 7H), 1.95 (m, 1H), 1.16 (d, $J$ = 7.3 Hz, 3H), 0.93 (d, $J$ = 6.9 Hz, 3H), 0.91 (d, $J$ = 6.9 Hz, 3H). HRMS calcd for $C_{26}H_{36}N_4O_7H^+$ [M + H]$^+$, 517.2657; found, 517.2658. HPLC gradient C, $t_R$ = 10.1 min (HPLC purity >95%, 230 nm).

*(Z)-6-((2R,5R,8R,11R)-8-((1H-Indol-3-yl)methyl)-2-isopropyl-5-methyl-3,6,9,13-tetraoxo-1,4,7,10-tetraazacyclotridecan-11-yl)hex-4-enoic Acid (**9c**).* The linear peptide **34c** (35 mg, 0.051 mmol) was dissolved in DMF (130 mL) and $i$-Pr$_2$NEt (71 μL, 0.41 mmol, 10 equiv) and HATU (29 mg, 0.77 mmol, 1.5 equiv) were added. The solution was stirred for 16 h, and the reaction mixture was concentrated in vacuo. Purification by VLC afforded the protected cyclic peptide (30 mg) as a white amorphous solid, which was used without further purification. UPLC-MS gradient B, $t_R$ = 2.52 min. MS calcd for $C_{35}H_{49}N_5O_8H^+$ [M + H]$^+$, 668.4; found, 668.4. LiOH (54 mg, 2.25 mmol, 50 equiv) in water (4 mL) was added to the crude cyclic peptide (30 mg, ∼0.045 mmol) in THF (4 mL). After 17 h, the THF was removed in vacuo and the aqueous phase was acidified to pH 1−2 with 2 M HCl and extracted with $CH_2Cl_2$ (3 × 50 mL) and EtOAc (2 × 40 mL). The combined organic phases were pooled and concentrated in vacuo to give the crude Boc-protected cyclic peptide **35** as orange oil. The crude cyclic peptide was dissolved in $CH_2Cl_2$ (6 mL), and TFA was added (6 mL). After 1 h of stirring, the reaction mixture was concentrated in vacuo and the resulting residue was dissolved in DMF−MeCN−water (1:2:2, 2.5 mL) and purified by preparative HPLC to afford **9c** (2.6 mg, 5%, 6 steps) as a white solid. Two conformations were observed in a 97:3 ratio determined by integration of $^1$H NMR peaks. Characterization is given for the major conformation. $^1$H NMR (500 MHz, DMSO-$d_6$) $\delta$ 12.08 (br s, 1H), 10.87 (br s, 1H), 7.83 (m, 1H), 7.51 (m, 3H), 7.33 (d, $J$ = 8.1 Hz, 1H), 7.27 (m, 1H), 7.08 (s, 1H), 7.06 (t, $J$ = 7.8 Hz, 1H), 6.97 (t, $J$ = 7.4 Hz, 1H), 5.47 (m, 1H), 5.33 (m, 1H), 4.36 (m, 1H), 4.30 (m, 1H), 3.92 (m, 1H), 3.82 (t, $J$ = 6.5 Hz, 1H), 3.01 (m, 2H), 2.45 (m, 1H), 2.33−2.14 (m, 7H), 1.94 (m, 1H), 1.15 (d, $J$ = 7.3 Hz, 3H), 0.92 (d, $J$ = 7.4 Hz, 3H), 0.90 (d, $J$ = 7.4 Hz, 3H). HRMS calcd for $C_{28}H_{37}N_5O_6H^+$ [M + H]$^+$, 540.2817; found, 540.2814. HPLC gradient C, $t_R$ = 10.5 min (HPLC purity >95%, 230 nm).

**Biochemical Profiling.** *HDAC Assay Materials.* HDAC1 (purity ≥62% by SDS-PAGE according to the supplier), HDAC2 (full length, purity ≥94% by SDS-PAGE according to the supplier), HDAC3−NCoR1 complex (purity ≥80% by SDS-PAGE according to supplier, HDAC4 (purity ≥60% by SDS-PAGE according to the supplier) and HDAC 5 (full length, purity ≥90% by SDS-PAGE according to the supplier), HDAC8 (purity ≥90% by SDS-PAGE according to the

supplier), HDAC9 (full length, purity ≥76% by SDS-PAGE according to the supplier) and HDAC10 (purity ≥21% by SDS-PAGE according to the supplier) were purchased from BPS Bioscience (San Diego, CA 92121). HDAC 7 (purity ≥90% by SDS-PAGE according to the supplier) was purchased from Millipore (Temecula, CA 92590). HDAC6 (purity ≥85% by SDS-PAGE according to the supplier) and HDAC11 (purity ≥50% by SDS-PAGE according to the supplier) were purchased from Enzo Life Sciences (Postfach, Switzerland). The HDAC assay buffer (50 mM tris/Cl, pH 8.0, 137 mM NaCl, 2.7 mM KCl, 1 mM $MgCl_2$) was added bovine serum albumin (0.5 mg/mL). Trypsin (10,000 units/mg, TPCK treated from bovine pancreas) was from Sigma-Aldrich (Steinheim, Germany). All peptides were purified to homogeneity (>95% purity by $HPLC_{230\ nm}$ using reversed-phase preparative HPLC), and the white fluffy materials obtained by lyophilization were kept at −20 °C. For assaying, peptide substrates were reconstituted in DMSO to give 5−10 mM stock solutions, the accurate concentration of which were determined by UV using the extinction coefficient for tyrosine and tryptophane at 280 nm; $\varepsilon$ = 1280 $M^{-1} \times cm^{-1}$ and $\varepsilon$ = 5690 $M^{-1} \times cm^{-1}$, respectively. Phenylalanine-containing compounds were adjusted after coinjection on HPLC with a Phe sample of known concentration.

*In Vitro Cell Cytotoxicity Assay.* Cell viability was assessed using the MTT cell growth kit (Millipore, Billerica, MA, USA). EB-3 cells (ATCC, Manassas, VA, USA) were seeded in 96-well flat-bottom cell culture plates (Nunc, Thermo Fischer Scientific, Roskilde, Denmark) with a density of $2 \times 10^4$ cells in 90 μL of culture media composed of RPMI-1640 (ATCC, Manassas, VA, USA) supplemented with 10% fetal calf serum (FCS) (Sera Scandia, Hellerup, Denmark) and 1% penicillin/streptomycin (Gibco, Invitrogen, Taastrup, Denmark). After cultivation overnight at 37 °C in humid 5% $CO_2$ atmosphere, 10 μL of culture medium containing appropriate concentrations of the five compounds or the vehicle (DMSO) as control were added. The 2-fold dilution series were prepared from DMSO stock solutions (5 mM for azumamide C and 10 mM for the remaining compounds) to give final assay concentrations starting from 50 μM for azumamide C, SAHA, and apicidin while 100 μM was applied for azumamide B and compound **9**. After incubation for 3 days, MTT solution was freshly prepared by dissolving 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT, 50 mg) in PBS (10 mL, both supplied in the MTT assay kit) and 10 μL of this solution was added to each well. The cells were incubated for another 4 h at 37 °C in humid 5% $CO_2$ atmosphere for development to take place. The purple formazan crystals produced were dissolved by addition of 2-propanol (100 μL/well) containing HCl (0.04 M), and the absorbance was measured at 570 nm with background subtraction at 630 nm on a Tecan Sunrise ELISA plate reader (Tecan, Switzerland). All assays were performed three times in duplicate.

*In Vitro Histone Deacetylase Inhibition Assays.* For inhibition of recombinant human HDACs, the dose−response experiments with SAHA as internal positive control in each plate were performed in black low binding NUNC 96-well microtiter plates. Dilution series (3-fold dilution, six concentrations) were prepared in HDAC assay buffer from 5−10 mM DMSO stock solutions. The appropriate dilution of inhibitor (5 μL of 5× the desired final concentration) was added to each well followed by HDAC assay buffer (10 μL) containing substrate [*Ac-Leu-Gly-Lys(Ac)-AMC* (20 μM) for HDAC1, 2, 3, 6, and 11; (200 μM) for HDAC8; and *Ac-Arg-His-Lys(Ac)-Lys(Ac)-AMC* (5 μM) for HDAC10]. Finally, a solution of the appropriate HDAC (10 μL) was added, and the plate was incubated at 37 °C for 30 min. Final HDAC concentrations: HDAC1, 6 ng/μL; HDAC2, 2 ng/μL; HDAC3−NCoR1, 0.2 ng/μL; HDAC6, 2.4 ng/μL; HDAC8, 0.2 ng/μL; HDAC10, 14 ng/μL; and HDAC11, 10 ng/μL. Then trypsin (25 μL, 0.4 mg/mL) was added, and the assay development was allowed to proceed for 15−30 min at room temperature before the plate was read using a PerkinElmer Enspire plate reader with excitation at 360 nm and detecting emission at 460 nm. Each assay was performed in duplicate. The data were analyzed by nonlinear regression using GraphPad Prism to afford $IC_{50}$ values from the dose−response experiments, and $K_i$ values were determined from the Cheng−Prusoff equation [$K_i = IC_{50}/(1 + [S]/K_m)$] assuming a standard fast-on−fast-off mechanism of inhibition.

*NMR Spectroscopy.* NMR spectra were acquired using standard pulse sequences on a Unity Inova 500 Varian spectrometer (499.9 MHz for $^1H$, 125.7 MHz for $^{13}C$) or a Bruker Avance 800 MHz spectrometer (798.9 MHz for $^1H$ and 200.9 MHz for $^{13}C$) located at the Danish Instrument Centre for NMR Spectroscopy of Biological Macromolecules at Carlsberg Laboratory. The deuterated solvent used for all compounds was DMSO-$d_6$. For typical homonuclear 2D experiments, 4096 data points were recorded in the direct dimension and 512 in the indirect dimension. Typical d1 times were from 2 to 4 s, which were found to give identical results compared to 7 s (>5 × T1 of the longest T1). The T1 times were investigated for **5a** and a model compound (**9c** attenuated at C6), and most nuclei were found to have T1 values up to approximately 1 s with a few aromatic protons exhibiting up to 1.3 s relaxation times. All *J*-couplings were extracted from the 1D $^1H$ and DQF-COSY spectra. Distances were obtained from 2D NOESY or ROESY experiments using the isolated spin pair approximation (ISPA),[39,40] and the linear range was increased by the method suggested by Macura et al.[50] The used mixing time was 150 ms for all compounds. Initially, different mixing times were applied to construct build-up curves to ensure the use of only cross-peaks that fit ISPA. The *J*-couplings from angles were calculated by the Karplus equation[51] as well as the HLA equation[52] for the β-peptide protons.

*Prime/MMGBSA Binding Energy Estimation.* We estimated binding free energies (dG_Bind) with Prime/MMGBSA (Prime version 3.3, Schrödinger, LLC, New York, NY, 2013) calculations on the Embrace-optimized complexes of HDAC3 and azumamide E analogues generated with the "mutation" approach (see the Supporting Information, Methods for details). The computational protocol was tested with the following settings for receptor and ligand flexibility: (a) frozen $Zn^{2+}$ and coordinating residues, flexible azumamide, and residues within 6 Å of azumamide, (b) flexible $Zn^{2+}$ and coordinating residues, flexible azumamide, and residues within 6 Å of azumamide, (c) frozen receptor and flexible azumamide, (d) frozen receptor and frozen ligand, and (e) repetition of the latter calculation without water and $K^+$. To ensure movement of the same residues across all complexes in calculations with receptor flexibility, the residues used throughout were those detected within a 6 Å radius of the substrate in the azumamide E:HDAC3 complex. Results from Prime/MMGBSA calculations are given in Supporting Information, Table S17.

*DFT.* The interaction between the $\beta^2$-substituent in azumamide E analogues and Phe200 in HDAC3B was further examined with density functional theory (DFT) calculations. Because the size of HDAC3−azumamide complexes precluded DFT calculations on the entire systems, we carried out the calculations on two truncated model systems referred to as "small" and "large" generated from each Embrace optimized complex. The small system consisted of the HDAC3B Phe200 side chain and the azumamide $\beta^2$-carbon and its substituents (hydrogen, methyl, or dimethyl). The large system consisted of the HDAC3B Phe200 side chain and the entire azumamide. For both types of systems, any bonds broken by the excision from the Embrace complexes were saturated with hydrogen. The interaction energy between the Phe200 side chain and azumamide $\beta^2$-substituent was calculated in Jaguar (version 8.1, Schrödinger, LLC, New York, NY, 2013) at the B3LYP-D3/cc-pvtz level of theory using the counterpoise (CP) correction for the basis set superposition error (BSSE).[53] To preserve Embrace geometries, no geometry optimization was employed during the counterpoise calculations.

## ■ ASSOCIATED CONTENT

### Ⓢ Supporting Information

Supplementary methods; supplementary references; chemistry; HDAC profiling; NMR structures; molecular dynamics simulations of macrocycles; docking to HDAC3 and HDAC crystal structure alignments; preliminary biochemical in vitro HDAC screening; NMR structural studies; computational studies; $^1H$ and $^{13}C$ NMR spectra; ROESY spectra; CIF files

for the X-ray crystal structures. This material is available free of charge via the Internet at http://pubs.acs.org.

## AUTHOR INFORMATION


### Corresponding Author

*Phone: (+45) 45252105. E-mail: cao@sund.ku.dk.

### Present Address

§For C.A.O.: Center for Biopharmaceuticals and Department of Drug Design & Pharmacology, University of Copenhagen, Universitetsparken 2, DK-2100, Copenhagen, Denmark.


### Author Contributions

#A.R.M. and J.S.V. contributed equally to this work.

A.R.M. and J.S.V. performed all syntheses and HDAC profiling assays and wrote an initial draft of the manuscript. C.H. and C.H.G. performed the NMR structure determination and contributed to manuscript preparation. N.J.C. and P.F. performed the molecular modeling and contributed to manuscript preparation. P.H. performed the small molecule X-ray crystallography. T.F. performed the cell-based assays. C.A.O. designed and supervised the study and wrote the paper with input from all the coauthors.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was supported by the Lundbeck Foundation (Young Group Leader Fellowship, C.A.O.), the Danish Independent Research Council–Natural Sciences (Steno grant no. 10-080907, C.A.O.), and the Carlsberg Foundation (equipment). The Danish Instrument Center for NMR Spectroscopy of Biological Macromolecules is acknowledged for access to their 800 MHz spectrometer. We thank Anne Hector (DTU Chemistry) for assistance with NMR spectroscopy, Tina Gustafsson (DTU Chemistry) for technical assistance with UPLC-MS, and Esin Güven (KBIG, SSI) for cell-based assays. Dr. A. S. Madsen and Helle M. E. Kristensen are gratefully acknowledged for assistance with the biochemical HDAC profiling.

## ABBREVIATIONS USED

AMC, 7-amino-4-methylcoumarin; Bcl-2, B-cell lymphoma 2; BSSE, basis set superposition error; CP, counterpoise; DFT, density functional theory; DIC, *N*,*N*′-diisopropylcarbodiimide; EBV, Epstein–Barr virus; ELISA, enzyme-linked immunosorbent assay; FCS, fetal calf serum; HATU, 2-(1*H*-azabenzotriazol-1-yl)-1,1,3,3,-tetramethyluronium hexafluorophosphate; HDAC, histone deacetylase; HIV, human immunodeficiency virus; HPLC, high-performance liquid chromatography; ISPA, isolated spin pair approximation; LDA, lithium diisopropylamide; MD, molecular dynamics; MS, mass spectrometry; MTT, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide; NAD, nicotinamide adenine dinucleotide; NCoR1, nuclear receptor corepressor 1; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; PBS, phosphate buffered saline; PMS, *N*-methyldibenzopyrazine methyl sulfate; RMSD, root-mean-square deviation; ROE, rotating frame nuclear Overhauser effect; RP-HPLC, reversed-phase high-performance liquid chromatography; SAR, structure–activity relationship; SPPS, solid-phase peptide synthesis; SPS, solid-phase synthesis; TFA, trifluoroacetic acid; TSA, trichostatin A; VLC, vacuum liquid chromatography

## REFERENCES


(1) Haberland, M.; Montgomery, R. L.; Olson, E. N. The Many Roles of Histone Deacetylases in Development and Physiology: Implications for Disease and Therapy. *Nature Rev. Genet.* **2009**, *10*, 32–42.

(2) Choudhary, C.; Kumar, C.; Gnad, F.; Nielsen, M. L.; Rehman, M.; Walther, T. C.; Olsen, J. V.; Mann, M. Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. *Science* **2009**, *325*, 834–840.

(3) Du, J.; Zhou, Y.; Su, X.; Yu, J. J.; Khan, S.; Jiang, H.; Kim, J.; Woo, J.; Kim, J. H.; Choi, B. H.; He, B.; Chen, W.; Zhang, S.; Cerione, R. A.; Auwerx, J.; Hao, Q.; Lin, H. Sirt5 is a NAD-Dependent Protein Lysine Demalonylase and Desuccinylase. *Science* **2011**, *334*, 806–809.

(4) Tan, M.; Luo, H.; Lee, S.; Jin, F.; Yang, J. S.; Montellier, E.; Buchou, T.; Cheng, Z.; Rousseaux, S.; Rajagopal, N.; Lu, Z.; Ye, Z.; Zhu, Q.; Wysocka, J.; Ye, Y.; Khochbin, S.; Ren, B.; Zhao, Y. Identification of 67 Histone Marks and Histone Lysine Crotonylation as a New Type of Histone Modification. *Cell* **2011**, *146*, 1016–1028.

(5) Madsen, A. S.; Olsen, C. A. Profiling of Substrates for Zinc-Dependent Lysine Deacylase Enzymes: HDAC3 Exhibits Decrotonylase Activity in Vitro. *Angew. Chem., Int. Ed.* **2012**, *51*, 9083–9087.

(6) Jiang, H.; Khan, S.; Wang, Y.; Charron, G.; He, B.; Sebastian, C.; Du, J. T.; Kim, R.; Ge, E.; Mostoslavsky, R.; Hang, H. C.; Hao, Q.; Lin, H. N. SIRT6 Regulates TNF-alpha Secretion Through Hydrolysis of Long-Chain Fatty Acyl Lysine. *Nature* **2013**, *496*, 110–113.

(7) Tan, M.; Peng, C.; Anderson, K. A.; Chhoy, P.; Xie, Z.; Dai, L.; Park, J.; Chen, Y.; Huang, H.; Zhang, Y.; Ro, J.; Wagner, G. R.; Green, M. F.; Madsen, A. S.; Schmiesing, J.; Peterson, B. S.; Xu, G.; Ilkayeva, O. R.; Muehlbauer, M. J.; Braulke, T.; Muhlhausen, C.; Backos, D. S.; Olsen, C. A.; McGuire, P. J.; Pletcher, S. D.; Lombard, D. B.; Hirschey, M. D.; Zhao, Y. Lysine Glutarylation Is a Protein Posttranslational Modification Regulated by SIRT5. *Cell Metab.* **2014**, *19*, 605–617.

(8) West, A. C.; Johnstone, R. W. New and Emerging HDAC Inhibitors for Cancer Treatment. *J. Clin. Invest.* **2014**, *124*, 30–39.

(9) Marks, P. A.; Breslow, R. Dimethyl Sulfoxide to Vorinostat: Development of this Histone Deacetylase Inhibitor as an Anticancer Drug. *Nature Biotechnol.* **2007**, *25*, 84–90.

(10) Prince, H. M.; Dickinson, M. Romidepsin for Cutaneous T-Cell Lymphoma. *Clin. Cancer Res.* **2012**, *18*, 3509–3515.

(11) Kazantsev, A. G.; Thompson, L. M. Therapeutic Application of Histone Deacetylase Inhibitors for Central Nervous System Disorders. *Nature Rev. Drug Discovery* **2008**, *7*, 854–868.

(12) Fischer, A.; Sananbenesi, F.; Wang, X.; Dobbin, M.; Tsai, L. H. Recovery of Learning and Memory is Associated with Chromatin Remodelling. *Nature* **2007**, *447*, 178–182.

(13) Archin, N. M.; Liberty, A. L.; Kashuba, A. D.; Choudhary, S. K.; Kuruc, J. D.; Crooks, A. M.; Parker, D. C.; Anderson, E. M.; Kearney, M. F.; Strain, M. C.; Richman, D. D.; Hudgens, M. G.; Bosch, R. J.; Coffin, J. M.; Eron, J. J.; Hazuda, D. J.; Margolis, D. M. Administration of Vorinostat Disrupts HIV-1 Latency in Patients on Antiretroviral Therapy. *Nature* **2012**, *487*, 482–485.

(14) Gregoretti, I. V.; Lee, Y. M.; Goodson, H. V. Molecular Evolution of the Histone Deacetylase Family: Functional Implications of Phylogenetic Analysis. *J. Mol. Biol.* **2004**, *338*, 17–31.

(15) Imai, S.-i.; Guarente, L. Ten Years of NAD-Dependent SIR2 Family Deacetylases: Implications for Metabolic Diseases. *Trends Pharmacol. Sci.* **2010**, *31*, 212–220.

(16) Wagner, F. F.; Weiwer, M.; Lewis, M. C.; Holson, E. B. Small Molecule Inhibitors of Zinc-Dependent Histone Deacetylases. *Neurotherapeutics* **2013**, *10*, 589–607.

(17) Newkirk, T. L.; Bowers, A. A.; Williams, R. M. Discovery, Biological Activity, Synthesis and Potential Therapeutic Utility of Naturally Occurring Histone Deacetylase Inhibitors. *Nat. Prod. Rep.* **2009**, *26*, 1293–1320.

(18) Sewald, N. J., H, D. *Peptides: Chemistry and Biology*, 2nd ed.; WILEY-VCH: Weinheim, 2009; p 578.

(19) Taunton, J.; Hassig, C. A.; Schreiber, S. L. A Mammalian Histone Deacetylase Related to the Yeast Transcriptional Regulator Rpd3p. *Science* **1996**, *272*, 408–411.

(20) Wang, D.; Helquist, P.; Wiest, O. Zinc Binding in HDAC Inhibitors: A DFT Study. *J. Org. Chem.* **2007**, *72*, 5446−5449.
(21) Singh, S. B.; Zink, D. L.; Polishook, J. D.; Dombrowski, A. W.; DarkinRattray, S. J.; Schmatz, D. M.; Goetz, M. A. Apicidins: Novel Cyclic Tetrapeptides as Coccidiostats and Antimalarial Agents from Fusarium Pallidoroseum. *Tetrahedron Lett.* **1996**, *37*, 8077−8080.
(22) Nakao, Y.; Yoshida, S.; Matsunaga, S.; Shindoh, N.; Terada, Y.; Nagai, K.; Yamashita, J. K.; Ganesan, A.; van Soest, R. W.; Fusetani, N.; Azumamides, A.-E. Histone Deacetylase Inhibitory Cyclic Tetrapeptides from the Marine Sponge Mycale Izuensis. *Angew. Chem., Int. Ed.* **2006**, *45*, 7553−7557.
(23) Maulucci, N.; Chini, M. G.; Micco, S. D.; Izzo, I.; Cafaro, E.; Russo, A.; Gallinari, P.; Paolini, C.; Nardi, M. C.; Casapullo, A.; Riccio, R.; Bifulco, G.; De Riccardis, F. Molecular Insights Into Azumamide E Histone Deacetylases Inhibitory Activity. *J. Am. Chem. Soc.* **2007**, *129*, 3007−3012.
(24) Villadsen, J. S.; Stephansen, H. M.; Maolanon, A. R.; Harris, P.; Olsen, C. A. Total Synthesis and Full Histone Deacetylase Inhibitory Profiling of Azumamides A−E as well as beta(2)-epi-Azumamide E and beta(3)-epi-Azumamide E. *J. Med. Chem.* **2013**, *56*, 6512−6520.
(25) Horne, W. S.; Olsen, C. A.; Beierle, J. M.; Montero, A.; Ghadiri, M. R. Probing the Bioactive Conformation of an Archetypal Natural Product HDAC Inhibitor with Conformationally Homogeneous Triazole-Modified Cyclic Tetrapeptides. *Angew. Chem., Int. Ed.* **2009**, *48*, 4718−4724.
(26) Furumai, R.; Komatsu, Y.; Nishino, N.; Khochbin, S.; Yoshida, M.; Horinouchi, S. Potent Histone Deacetylase Inhibitors Built from Trichostatin A and Cyclic Tetrapeptide Antibiotics Including Trapoxin. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 87−92.
(27) Olsen, C. A.; Ghadiri, M. R. Discovery of Potent and Selective Histone Deacetylase Inhibitors via Focused Combinatorial Libraries of Cyclic Alpha3Beta-Tetrapeptides. *J. Med. Chem.* **2009**, *52*, 7836−7846.
(28) Taori, K.; Paul, V. J.; Luesch, H. Structure and Activity of Largazole, a Potent Antiproliferative Agent from the Floridian Marine Syanobacterium Symploca Sp. *J. Am. Chem. Soc.* **2008**, *130*, 1806−1807.
(29) Bowers, A. A.; Greshock, T. J.; West, N.; Estiu, G.; Schreiber, S. L.; Wiest, O.; Williams, R. M.; Bradner, J. E. Synthesis and Conformation−Activity Relationships of the Peptide Isosteres of FK228 and Largazole. *J. Am. Chem. Soc.* **2009**, *131*, 2900−2005.
(30) Michrowska, A.; Bujok, R.; Harutyunyan, S.; Sashuk, V.; Dolgonos, G.; Grela, K. Nitro-Substituted Hoveyda−Grubbs Ruthenium Carbenes: Enhancement of Catalyst Activity Through Electronic Activation. *J. Am. Chem. Soc.* **2004**, *126*, 9318−9325.
(31) Seiser, T.; Kamena, F.; Cramer, N. Synthesis and Biological Activity of Largazole and Derivatives. *Angew. Chem., Int. Ed.* **2008**, *47*, 6483−6485.
(32) Montero, A.; Beierle, J. M.; Olsen, C. A.; Ghadiri, M. R. Design, Synthesis, Biological Evaluation, and Structural Characterization of Potent Histone Deacetylase Inhibitors Based on Cyclic alpha/beta-Tetrapeptide Architectures. *J. Am. Chem. Soc.* **2009**, *131*, 3033−3041.
(33) Villadsen, J.; Kitir, B.; Wich, K.; Friis, T.; Madsen, A. S.; Olsen, C. A. An Azumamide C Analogue Without the Zinc-Binding Carboxylate Group. *Med. Chem. Commun.* **2014**, DOI: 10.1039/c4md00252k.
(34) Bradner, J. E.; West, N.; Grachan, M. L.; Greenberg, E. F.; Haggarty, S. J.; Warnow, T.; Mazitschek, R. Chemical Phylogenetics of Histone Deacetylases. *Nature Chem. Biol.* **2010**, *6*, 238−243.
(35) Kessler, H. Peptide Conformations 0.19. Conformation and Biological Activity of Cyclic-Peptides. *Angew. Chem., Int. Ed.* **1982**, *21*, 512−523.
(36) Shute, R. E.; Kawai, M.; Rich, D. H. Conformationally Constrained Biologically Active Peptides—Tentative Identification of the Antimitogenic Bioactive Conformer of the Naturally-Occurring Cyclic Tetrapeptides. *Tetrahedron* **1988**, *44*, 685−695.
(37) Izzo, I.; Maulucci, N.; Bifulco, G.; De Riccardis, F. Total Synthesis of Azumamides A and E. *Angew. Chem., Int. Ed.* **2006**, *45*, 7557−7560.
(38) Demmer, O.; Frank, A. O.; Hagn, F.; Schottelius, M.; Marinelli, L.; Cosconati, S.; Brack-Werner, R.; Kremb, S.; Wester, H. J.; Kessler, H. A Conformationally Frozen Peptoid Boosts CXCR4 Affinity and Anti-HIV Activity. *Angew. Chem., Int. Ed.* **2012**, *51*, 8110−8113.
(39) Bothnerby, A. A.; Stephens, R. L.; Lee, J. M.; Warren, C. D.; Jeanloz, R. W. Structure Determination of a Tetrasaccharide - Transient Nuclear Overhauser Effects in the Rotating Frame. *J. Am. Chem. Soc.* **1984**, *106*, 811−813.
(40) Kessler, H.; Griesinger, C.; Kerssebaum, R.; Wagner, K.; Ernst, R. R. Separation of Cross-Relaxation and J-Cross-Peaks in 2d Rotating-Frame NMR-Spectroscopy. *J. Am. Chem. Soc.* **1987**, *109*, 607−609.
(41) Watson, P. J.; Fairall, L.; Santos, G. M.; Schwabe, J. W. Structure of HDAC3 Bound to Co-Repressor and Inositol Tetraphosphate. *Nature* **2012**, *481*, 335−340.
(42) Bantscheff, M.; Hopf, C.; Savitski, M. M.; Dittmann, A.; Grandi, P.; Michon, A. M.; Schlegl, J.; Abraham, Y.; Becher, I.; Bergamini, G.; Boesche, M.; Delling, M.; Dumpelfeld, B.; Eberhard, D.; Huthmacher, C.; Mathieson, T.; Poeckel, D.; Reader, V.; Strunk, K.; Sweetman, G.; Kruse, U.; Neubauer, G.; Ramsden, N. G.; Drewes, G. Chemoproteomics Profiling of HDAC Inhibitors Reveals Selective Targeting of HDAC Complexes. *Nature Biotechnol.* **2011**, *29*, 255−265.
(43) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Fujii, A. Magnitude and Directionality of the Interaction Energy of the Aliphatic CH/pi Interaction: Significant Difference From Hydrogen Bond. *J. Phys. Chem. A* **2006**, *110*, 10163−10168.
(44) Plevin, M. J.; Bryce, D. L.; Boisbouvier, J. Direct Detection of CH/pi Interactions in Proteins. *Nature Chem.* **2010**, *2*, 466−471.
(45) Cole, K. E.; Dowling, D. P.; Boone, M. A.; Phillips, A. J.; Christianson, D. W. Structural Basis of the Antiproliferative Activity of Largazole, a Depsipeptide Inhibitor of the Histone Deacetylases. *J. Am. Chem. Soc.* **2011**, *133*, 12474−12478.
(46) Whitehead, L.; Dobler, M. R.; Radetich, B.; Zhu, Y. Y.; Atadja, P. W.; Claiborne, T.; Grob, J. E.; McRiner, A.; Pancost, M. R.; Patnaik, A.; Shao, W. L.; Shultz, M.; Tichkule, R.; Tommasi, R. A.; Vash, B.; Wang, P.; Stams, T. Human HDAC Isoform Selectivity Achieved via Exploitation of the Acetate Release Channel With Structurally Unique Small Molecule Inhibitors. *Bioorg. Med. Chem.* **2011**, *19*, 4626−4634.
(47) Richter-Larrea, J. A.; Robles, E. F.; Fresquet, V.; Beltran, E.; Rullan, A. J.; Agirre, X.; Calasanz, M. J.; Panizo, C.; Richter, J. A.; Hernandez, J. M.; Roman-Gomez, J.; Prosper, F.; Martinez-Climent, J. A. Reversion of Epigenetically Mediated BIM Silencing Overcomes chemoresistance in Burkitt lymphoma. *Blood* **2010**, *116*, 2531−2542.
(48) Leung, C. S.; Leung, S. S. F.; Tirado-Rives, J.; Jorgensen, W. L. Methyl Effects on Protein−Ligand Binding. *J. Med. Chem.* **2012**, *55*, 4489−2500.
(49) Bissantz, C.; Kuhn, B.; Stahl, M. A Medicinal Chemist's Guide to Molecular Interactions. *J. Med. Chem.* **2010**, *53*, 5061−5084.
(50) Macura, S.; Farmer, B. T.; Brown, L. R. An Improved Method for the Determination of Cross-Relaxation Rates from Noe Data. *J. Magn. Reson.* **1986**, *70*, 493−499.
(51) Pardi, A.; Billeter, M.; Wuthrich, K. Calibration of the Angular Dependence of the Amide Proton-$C^{\alpha}$ Proton Coupling Constants, $^{3}J_{HN\alpha}$, in a Globular Protein: Use of $^{3}J_{HN\alpha}$ for Identification of Helical Secondary Structure. *J. Mol. Biol.* **1984**, *180*, 741−751.
(52) Haasnoot, C. A. G.; DeLeeuw, F. A. A. M.; Altona, C. The Relationship Between Proton−Proton NMR Coupling Constants and Substituent Electronegativities, I. An Empirical Generalization of the Karplus Equation. *Tetrahedron* **1980**, *36*, 2783−2792.
(53) Boys, S. F.; Bernardi, F. Calculation of Small Molecular Interactions by Differences of Separate Total Energies—Some Procedures with Reduced Errors. *Mol. Phys.* **1970**, *19*, 553−566.

*Molecules* **2014**, *19*, 10898-10921; doi:10.3390/molecules190810898

OPEN ACCESS

*molecules*

**ISSN 1420-3049**

www.mdpi.com/journal/molecules

*Article*

# Dereplication Guided Discovery of Secondary Metabolites of Mixed Biosynthetic Origin from *Aspergillus aculeatus*

**Lene M. Petersen [1], Casper Hoeck [2], Jens C. Frisvad [1], Charlotte H. Gotfredsen [2] and Thomas O. Larsen [1,*]**

[1] Chemodiversity Group, Department of Systems Biology, Søltofts Plads B221, Technical University of Denmark, Kgs. Lyngby DK-2800, Denmark; E-Mails: lmape@bio.dtu.dk (L.M.P.); jcf@bio.dtu.dk (J.C.F.)

[2] Department of Chemistry, Kemitorvet B201, Technical University of Denmark, Kgs. Lyngby DK-2800, Denmark; E-Mails: casho@kemi.dtu.dk (C.H.); chg@kemi.dtu.dk (C.H.G.)

* Author to whom correspondence should be addressed; E-Mail: tol@bio.dtu.dk; Tel.: +45-2425-2632; Fax: +45-4588-4922.

*Received: 16 June 2014; in revised form: 15 July 2014 / Accepted: 16 July 2014 / Published: 25 July 2014*

---

**Abstract:** Investigation of the chemical profile of the industrially important black filamentous fungus *Aspergillus aculeatus* by UHPLC-DAD-HRMS and subsequent dereplication has led to the discovery of several novel compounds. Isolation and extensive 1D and 2D NMR spectroscopic analyses allowed for structural elucidation of a dioxomorpholine, a unique okaramine, an aflavinine and three novel structures of mixed biosynthetic origin, which we have named aculenes A–C. Moreover, known analogues of calbistrins, okaramines and secalonic acids were detected. All novel compounds were tested for antifungal activity against *Candida albicans*, however all showed only weak or no activity. *Aspergillus aculeatus* IBT 21030 was additionally shown to be capable of producing sclerotia. Examination of the sclerotia revealed a highly regulated production of metabolites in these morphological structures.

**Keywords:** *Aspergillus aculeatus*; Aspergilli; natural products; secondary metabolism; dereplication; sclerotia

---

## 1. Introduction

*Aspergillus aculeatus* is a filamentous fungus belonging to *Aspergillus* section *Nigri*—the black aspergilli. At least 145 metabolites have been characterized from the black aspergilli [1], many of which are biologically active. Naphtho-γ-pyrones (NGPs) such as aurasperone A and rubrofusarin B from *A. niger* are known to be antifungal [2], while other NGPs are reported to have antitumor activities [3]. Mycotoxins produced by *A. niger*, such as ochratoxin A [4] or the fumonisins [5] are also known.

The black aspergilli can be divided into different clades. *A. aculeatus* belongs to the uniseriate black aspergilli and is closely related to *A. aculeatinus*, *A. uvarum*, *A. japonicus*, *A. fijiensis*, *A. trinidadensis*, *A. floridensis*, *A. brunneoviolaceus* and *A. violaceofuscus* [6,7]. These fungi differ from the other black aspergilli in their morphology, physiological behavior and in the production of secondary metabolites. The fungi belonging to this group can produce secondary metabolites, which can be both polyketide (PK) and nonribosomal peptide (NRP) derived or of mixed biosynthetic origin [1]. Several metabolites, including aculeacins A–G [8,9], CJ-15,183 [10], secalonic acids D and F [11] and okaramines H and I [12], have been reported in fungi identified as *A. aculeatus*. Furthermore, asperaculin A [13], aspergillusol A [14] and aculeatusquinones A–D [15] have been reported in marine strains of *A. aculeatus*. Different biological activities are reported for these metabolites, including antifungal (aculeacins A-G and CJ-15,183), enzyme inhibitory (CJ-15,183 and aspergilluson A), antimicrobial activities (secalonic D and F) and cytotoxicity (aculeatusquinone B and D).

While some black aspergilli are important in the biotechnological industry for production of enzymes and organic acids [16,17], some species can also be food and feed contaminants [18]. In fact, the black aspergilli are among the most common fungi connected to postharvest decay of fruit, beans and nuts, and *A. aculeatus* is no exception [18,19].

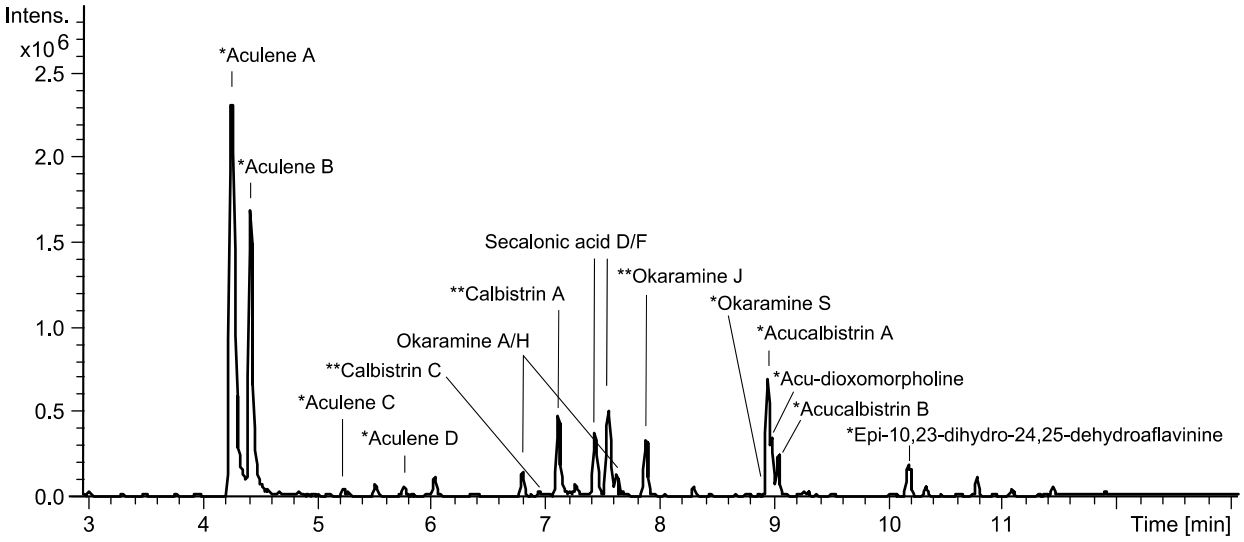
*A. aculeatus* is used to produce important industrial enzymes such as cellulases [20], xylanases [21,22] and proteases [23], which are used commercially in the food and feed industries. Moreover a marine strain of *A. aculeatus* has tested active against *Staphylococcus aureus* [24]. Based on both the industrial applications of *A. aculeatus* as well as food and feed contamination hazards it is of great importance to know what metabolites are being produced by this fungus. The aim of the current work has been to investigate the chemical profile of *A. aculeatus* seeking to discover novel compounds and to test for the antifungal activity of its metabolites.

## 2. Results and Discussion

Initial analysis of the chemistry of *A. aculeatus* involved two strains (IBT 21030 and IBT 3244) which were investigated on a series of solid media (YES, CYA, MEA, OAT and CREA) [25]. The strains were cultivated at 25 and 30 °C in the dark for 7 days and were investigated with micro-scale extractions [26]. The secondary metabolite profiles were analyzed with UHPLC-DAD-HRMS, which demonstrated that the strains showed highly similar chemical profiles. However, more metabolites were produced by the *A. aculeatus* IBT 21030 strain, hence this strain was chosen for further work. The medium used in the cultivations had varying effect on metabolite production. While cultivation on MEA, OAT and CREA did not lead to diverse production, cultivation on both CYA and YES lead to production of several different metabolites. The YES medium was selected for further work, and it was

found that the optimal cultivation temperature was 25 °C. The *A. aculeatus* strain was inoculated on YES media and incubated at 25 °C and analyzed after 2, 4, 7, and 10 days to test the effect of incubation time on the metabolite production. These experiments showed that 7 days of incubation gave the optimum production of secondary metabolites. Based on these initial experiments, *A. aculeatus* was inoculated on 200 plates of YES and incubated at 25 °C for seven days in the dark. The results of the dereplication as well as isolation and elucidation of novel compounds are depicted in Figure 1.

**Figure 1.** Base peak chromatogram illustrating the dereplication of some of the major compounds in the extract from *A. aculeatus* IBT 21030 as well as the results of purification and structural elucidation of novel compounds. Based on cultivation on YES media for 7 days at 25 °C in the dark. * Novel compounds reported here for the first time. ** Compounds reported from *A. aculeatus* for the first time.

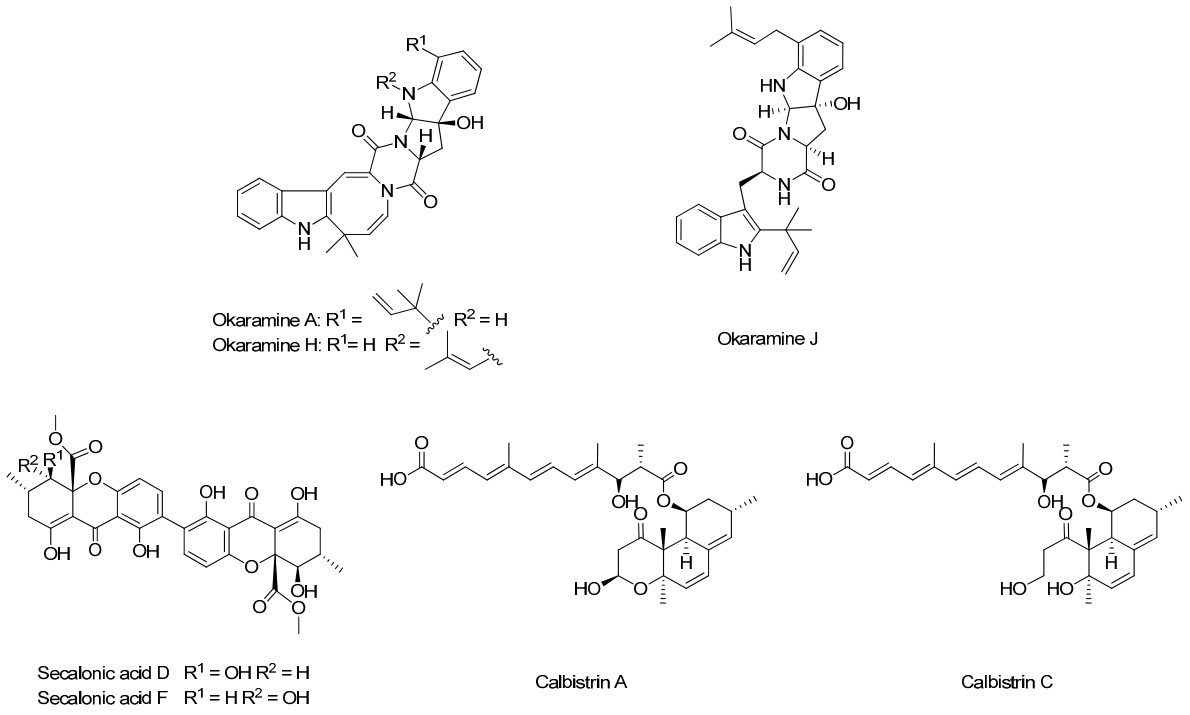



Dereplication was performed by UV- and HRMS-data to identify known compounds. An in-house library of microbial metabolites [27] as well as the AntiBase 2012 natural products database [28] were used for identification. Many of the major peaks could be dereplicated as compounds already known to be produced by *A. aculeatus* (Figures 1 and 2). This included secalonic acids D and F [11], as well as okaramines A and H [12]. One further okaramine was produced, which could not be unambiguously dereplicated, since the HRMS and UV data pointed towards either okaramine C, J *or* L. The compound was isolated and the structure was elucidated by 1D and 2D NMR and the structure was established to correspond to okaramine J, which has previously been described from *Penicillium ochrochloron* (formerly identified as *Penicillium simplicissimum*) [29,30].

Two calbistrin analogues could also be dereplicated from *A. aculeatus*. Calbistrins A–D, have been described from *Penicillium restrictum* by Jackson *et al.* [31] and the structures of the four compounds were later elucidated by the same group [32]. Calbistrin A is biologically active as an antifungal agent, a promoter of nerve growth factor production and a cholesterol lowering agent. Identification of the two calbistrins was not possible solely based on HRMS and DAD data, as calbistrins pairwise have the same molecular formula (A and B have the same molecular formula and likewise for calbistrins C and D). The

two compounds were therefore isolated and their structures elucidated by 1D and 2D NMR, and it was shown that the two compounds produced by *A. aculeatus* were calbistrin A and C, Figure 2.

**Figure 2.** Known compounds dereplicated from *A. aculeatus*. Okaramines A and H as well as secalonic acids have previously been reported from the organism, whereas the calbistrins and okaramine J are reported from *A. aculeatus* for the first time.




*2.1. A Novel Class of Compounds, Named the aculenes, Has a Mixed Biosynthetic Origin*

The two major metabolites produced by *A. aculeatus*, aculenes A and B (Figure 1) have previously been detected by UHPLC-DAD-HRMS in the related black aspergillus *A. saccharolyticus* [33], but the structures have not been elucidated. The molecular formulae of the two compounds were determined by HRMS to be $C_{19}H_{25}NO_3$ and $C_{19}H_{27}NO_3$, respectively. The mass difference of 2 Da and the fact that the two compounds eluted very close to each other suggested that the compounds were related and only differed by a double bond. The UV-spectra differed greatly as depicted in Figure 3, indeed indicating aculene A to contain a more conjugated chromophore.

The structures of the two compounds were elucidated by 1D- and 2D NMR spectroscopy and the structures of aculenes A and B are depicted in Figure 4. The conjugation in aculene A was in agreement with the recorded UV spectrum. The only difference with aculene B is the absent double bond between C-11 and C-12 and as seen, this structure was also in accordance with the obtained UV spectrum. The $^1$H-NMR spectra of both aculenes A and B revealed the presence of one $H_\alpha$ proton (H-5), which was found to be in a COSY spin system with three methylene groups (H-2, H-3 and H-4), which where elucidated as a proline ring. The remaining resonances in the $^1$H-NMR spectra belonged to a fused five and seven membered ring system. The linking between the COSY spin systems for this part and assignments of the quaternary carbons were accomplished through detailed analysis of the

HMBC experimental data. Some of the important HMBC correlations for the elucidation of aculene B are depicted in Figure 5.

**Figure 3.** UV-spectra of selected *A. aculeatus* metabolites. The remaining UV spectra can be found in the Supporting Information.

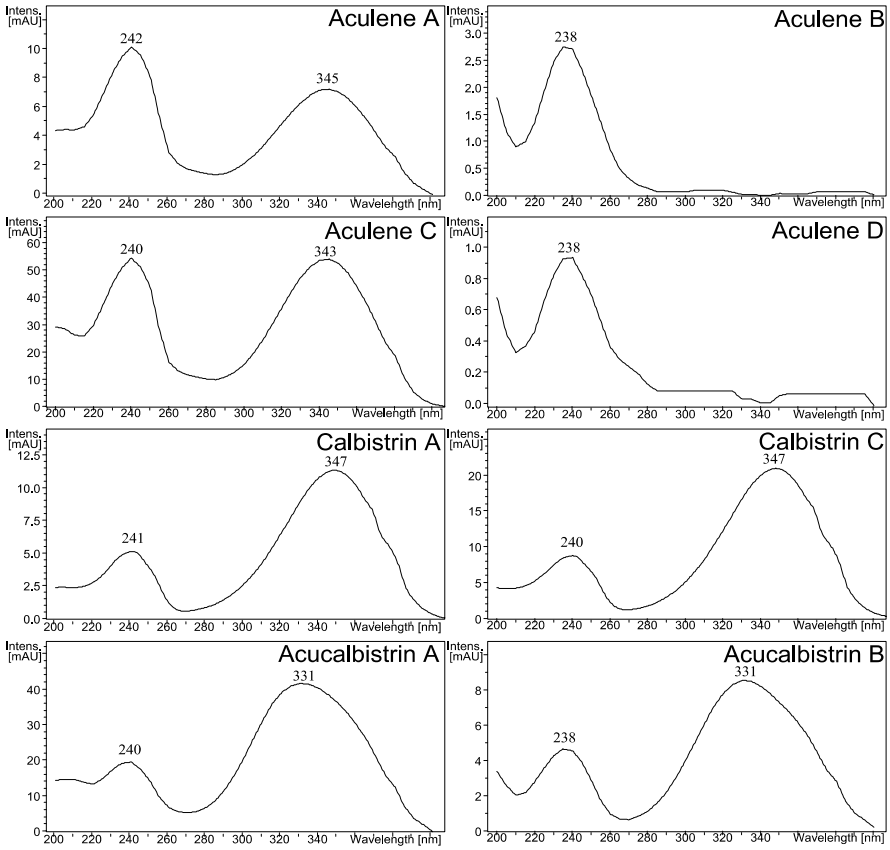
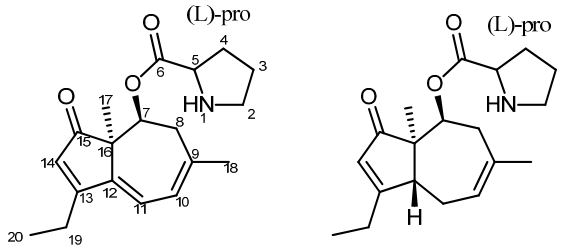
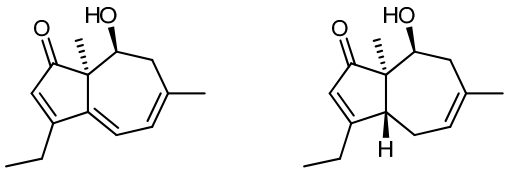



**Figure 4.** The structure of aculenes A–D. The structures of aculenes A–C have been verified by 1D and 2D NMR spectroscopy, whereas the structure of aculene D has been suggested based on UHPLC-DAD-HRMS data. The stereochemistry shown is relative.




Aculene A Aculene B




Aculene C Aculene D

**Figure 5. A**. Important HMBC correlations in the elucidation of aculene B. **B**. The NOEs used to solve the stereochemistry. Similar HMBCs and NOEs were observed for aculenes A and C, see Table 1 and Supplementary Table S2.

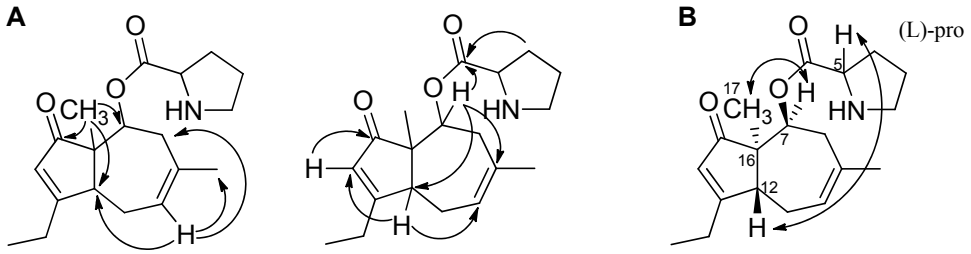



**Table 1.** $^{1}$H-NMR spectroscopic data (500 and 800 MHz, DMSO-$d_6$) and HMBCs for aculenes A–C.

| | Aculene A | | Aculene B | | Aculene C | |
|---|---|---|---|---|---|---|
| **Position** | **$\delta_H$ (*J* in Hz)** | **HMBC** | **$\delta_H$ (*J* in Hz)** | **HMBC** | **$\delta_H$ (*J* in Hz)** | **HMBC** |
| **1** | - | - | - | - | | |
| **2** | 3.15 (m) | - | 3.21 (m) | 3,4,5 | | |
| **2'** | 3.09 (m) | - | 3.14 (m) | 3,4,5 | | |
| **3** | 1.83 (m) | - | 1.84 (m) | 2,4,5 | | |
| **3'** | 1.67 (m) | - | 1.76 (m) | 2,4,5 | | |
| **4** | 1.94 (m) | 2,6 | 2.05 (m) | 2,3,5,6 | | |
| **4'** | 1.46 (m) | 2,6 | 1.55 (m) | 2,6 | | |
| **5** | 4.23 (t 7.7) | 6 | 4.31 (t 7.8) | 2,3,4,6 | | |
| **6** | - | - | - | - | | |
| **7** | 5.33 (dd 4.4, 2.4) | - | 5.28 (dd 5.3, 2.3) | 6,9,12,16 | 3.95 (t 3.5) | 9,13,15 |
| **8** | 2.82(br.d 20.8) | - | 2.67 (br. d 19.1) | 16 | 2.52 (br. d 20.1) | |
| **8'** | 2.55 (m) | 9,10,16 | 2.37 (m) | 7,9,10,16 | 2.31 (m) | 7,9,10,16 |
| **9** | - | - | - | - | - | - |
| **10** | 5.98 (d 7.4) | 8,12,18 | 5.56 (m) | 8,11,12,18 | 5.87 (d 7.3) | 8,12,18 |
| **11** | 6.24 (d 7.4) | 9,13,16 | 2.12 (m) | - | 6.08 (d 7.4) | 9,13,16 |
| **11'** | - | - | 2.52 (m) | 18 | - | - |
| **12** | - | - | 3.31 (m) | 10,11,14 | - | - |
| **13** | - | - | - | - | - | - |
| **14** | 6.03 (s) | 12,13,15,16,19 | 5.81 (q 1.8) | 12,13,15,16,19 | 5.93 (s) | 12,13,15,16, 19 |
| **15** | - | - | - | - | - | - |
| **16** | - | - | - | - | - | - |
| **17** | 0.96 (s) | 7,12,15,16 | 0.96 (s) | 7,12,15,16 | 0.79 (s) | 7,12,15,16 |
| **18** | 1.85 (s) | 8,9,10 | 1.68 (s) | 8,9,10 | 1.82 (s) | 8, 9,10 |
| **19** | 2.55 (m) | 13,20 | 2.39 (m) | 13,14,20 | 2.48 (m) | 12,13,14,20 |
| **20** | 1.16 (t 7.4) | 13,19 | 1.11 (t 7.3) | 13,19 | 1.14 (t 7.4) | 13,19 |
| **-OH** | - | - | - | - | 4.43 (s) | - |

The NMR data for aculene A were very comparable to those of aculene B (Tables 1 and 2), yet differences were observed corresponding to the additional double bond in aculene A. The chemical shift of H-10 and H-14 is seen to differ slightly and furthermore an additional resonance was present in the $^{1}$H-NMR spectrum in the aromatic/alkene area for compound A. This was located at a chemical

shift of $\delta_H$ = 6.24 ppm and assigned H-11. The only other location where the spectrum differed significantly was due to the extra diastereotopic $CH_2$ groups in the 7-membered ring of aculene B. For this, and the diastereotopic $CH_2$ group H8/H8' present in both compounds, rather large coupling constants were observed due to the geminal coupling. In compound A the H-8 doublet had a coupling constant of 20.8 Hz. Another notable chemical shift is the carbon chemical shift of C-13. This was at $\delta_C$ = 175.3 and 183.8 ppm, respectively for the two compounds, which is rather far downfield.

**Table 2.** $^{13}$C-NMR spectroscopic data (125 and 200 MHz, DMSO-$d_6$) for aculenes A–C.

| Position | Aculene A | Aculene B | Aculene C |
|---|---|---|---|
| | $\delta_C$ | $\delta_C$ | $\delta_C$ |
| **1** | - | - | - |
| **2** | 45.1 | 44.7 | - |
| **3** | 22.3 | 22.3 | - |
| **4** | 27.6 | 27.6 | - |
| **5** | 58.4 | 58.3 | - |
| **6** | 168.0 | 167.8 | - |
| **7** | 72.9 | 73.5 | 67.6 |
| **8** | 36.5 | 35.5 | 40.2 |
| **9** | 141.7 | 130.7 | 142.5 |
| **10** | 120.0 | 122.7 | 119.7 |
| **11** | 118.5 | 24.9 | 117.8 |
| **12** | 142.5 | 44.6 | 144.3 |
| **13** | 175.3 | 183.8 | 174.3 |
| **14** | 125.3 | 122.8 | 125.6 |
| **15** | 205.8 | 208.2 | 208.2 |
| **16** | 51.7 | 54.4 | 53.6 |
| **17** | 17.9 | 17.1 | 18.3 |
| **18** | 26.9 | 28.3 | 27.1 |
| **19** | 20.1 | 23.1 | 19.8 |
| **20** | 12.0 | 10.8 | 11.8 |

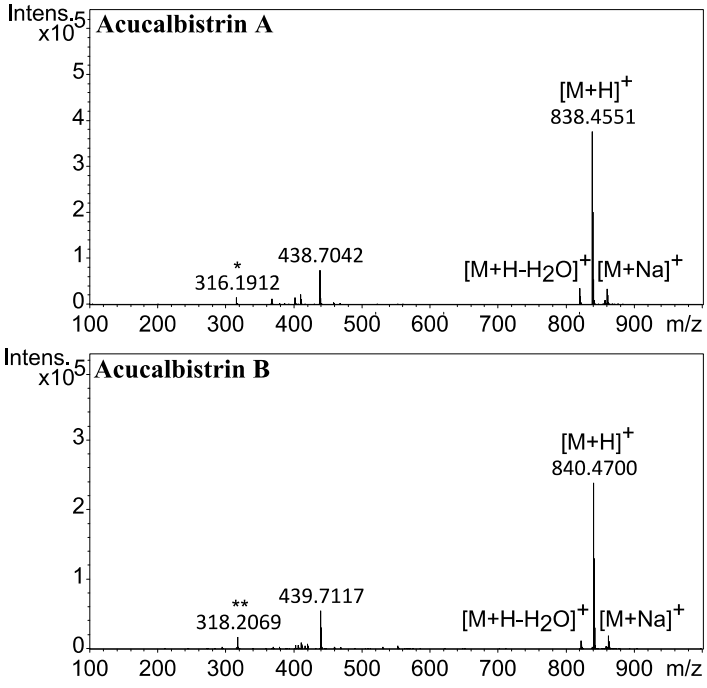
A third and related compound (aculene C) was also characterized. The UV spectrum was similar to the UV spectrum observed for aculene A (Figure 3) and the molecular formula was determined by HRMS to be $C_{14}H_{18}O_2$. The NMR data was comparable to that of aculene A (Tables 1 and 2), though with an absence of resonances originating from the proline part. Elucidation of the structure revealed that the compound was related to the aculenes, therefore it was named aculene C (Figure 4). Aculene C is a likely precursor to aculene A, having the same carbon skeleton, but missing the proline moiety. Analysis of the UHPLC-DAD-HRMS data showed a fourth compound, eluting close by aculene C, present in minute amounts, with a mass difference of 2 Da compared to aculene C, and with a UV spectrum very similar to aculene B (Figure 3). The HRMS data displayed the same adducts and fragmentation pattern as observed for aculene C. This compound, aculene D, is believed to have a similar structure as aculene C, but with the absence of the double bond between C-11 and C-12, as seen with the relation between aculenes A and B. This structure has not been verified by NMR experiments.

The stereochemistry of aculenes A-C was elucidated by NOEs, *J*-couplings and Marfey's reaction [34], from which it was determined that aculenes A and B contained L-proline. The stereocenters (C-5, C-7 and C-16) of aculenes A and B could not be connected, due to lack of NOEs from the proline to the bicyclic system and the free rotation of proline. The relative stereochemistry of stereocenters C-7 and C-16, as well as C-12 for aculene B, is depicted in Figure 4, suggested by NOEs and backcalculated $^{3}J$-couplings of H-7 (see Supplementary Figure S37 and Supplementary Tables S1 and S2). Relevant NOEs are depicted for aculene B in Figure 5.

We hypothesize the biosynthesis of the aculenes to originate from a terpene pathway. Aculene C and D consist of fourteen carbons, which could originate from a sesquiterpene with the loss of one carbon atom, possibly at C9. Aculene A and B furthermore contain the amino acid proline.

The two compounds marked as acucalbistrin A and B in Figure 1 were also targeted. These compounds were intriguing because they could not be dereplicated, and because the HRMS data suggested that the size were somewhat larger than typical secondary metabolites. The molecular formulas for the two compounds were determined to be $C_{50}H_{63}NO_{10}$ and $C_{50}H_{65}NO_{10}$, the mass spectra are depicted in Figure 6.

**Figure 6.** Mass spectra of acucalbistrin A (top) and acucalbistrin B (bottom). A fragment corresponding to *m/z* of the molecular ion $[M+H]^+$ of aculene A (*) and aculene B (**) are observed for acucalbistrin A and B respectively.




A feature noted in both mass spectra was a fragment corresponding to the mass of aculenes A and aculene B, respectively. Different chromatographic approaches were taken to purify both compounds, which however proved challenging as the compounds were unstable even under mild conditions. Apparently acucalbistrin A was degrading to aculene A and another compound X, while acucalbistrin B was degrading to aculene B and the same compound X. Analysis by UHPLC-DAD-HRMS of a small semipreparative proportion of this degradation product revealed to our surprise this compound to

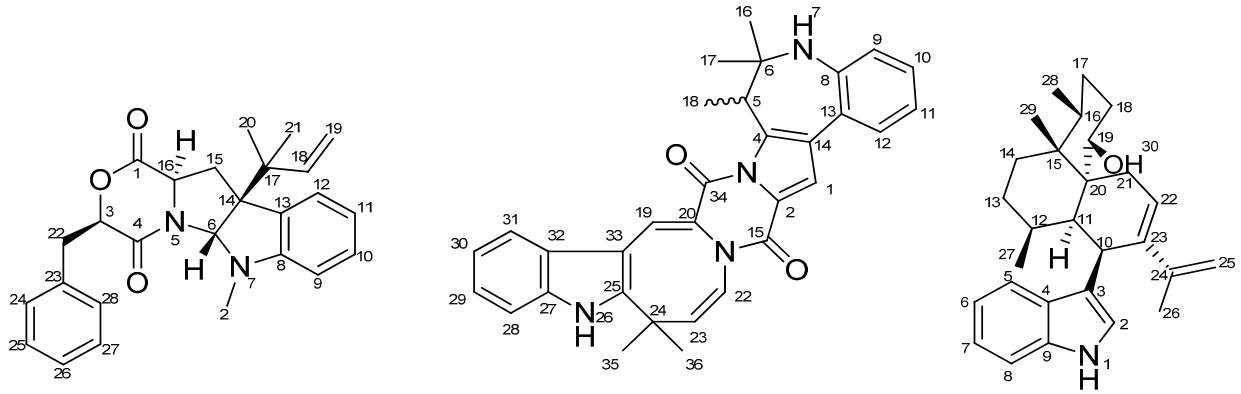
be calbistrin A. This was verified by the UHPLC-DAD-HRMS data by comparison to the known compounds, see Supplementary Figure S39.

The combined mass of aculene A and calbistrin A is higher than that of acucalbistrin A (and the same for the case of acucalbistrin B), where the difference corresponds to a single water molecule. We therefore hypothesize that the acucalbistrins undergo hydrolysis and with knowledge on the structures of aculenes A–B and calbistrin A, a possible structure could therefore be an amide bond between the nitrogen in the proline moiety in aculene to the carbonyl in the carboxylic acid part of calbistrin A (see Supplementary Figure S38). This would account for the observed masses. These acucalbistrin structures have not been verified by NMR spectroscopy. We note that the UV spectra of acucalbistrin A and B have $UV_{max}$ values comparable to those observed for calbistrin A (Figure 3). The slight shift observed may be due to the aculene moiety being attached directly to the carboxylic acid, which is part of the conjugated system in calbistrin A. The proposed structures are given in Supplementary Figure S38.

### *2.2. Discovery of Novel Indole Terpenoids*

Three other compounds were isolated and structure elucidated resulting in the novel compounds acu-dioxomorpholine, okaramine S and epi-10,23-dihydro-24,25-dehydroaflavinine with the structures shown in Figure 7.

**Figure 7.** The structure of acu-dioxomorpholine, okaramine S and epi-10,23-dihydro-24,25-dehydroaflavinine.




NMR data for the compounds are presented in Table 3, Table 4, and Table 5, respectively. The structure of acu-dioxomorpholine reminds one of a diketopiperazine, but this structure differs in that it contains a lactone. The chemical shift of the proton adjacent to the lactone oxygen (H-3) clearly differs compared to the chemical shift of a proton in ordinary diketopiperazines, which is typically reported to be around $\delta_H$ = 3.50–4.50 ppm [35,36]. For acu-dioxomorpholine the proton chemical shift is $\delta_H$ = 5.32 ppm. The carbon chemical shift of C-3 is also shifted significantly downfield, with $\delta_C$ = 78.1 ppm for the acu-dioxomorpholine, compared to around 60 ppm for diketopiperazines. Two compounds have been reported by Wang *et al.* with the same lactone feature [37], but differ by having leucine

incorporated instead of phenylalanine and by lacking the *N*-methylation. Otherwise the NMR data are comparable.

NOESY experiments enabled determination of the relative stereochemistry of acu-dioxomorpholine. NOE connectivities were found between H-3 and H-16 placing these protons on the same side of the ring system. These had no NOE connectivities to H-6, which however displayed NOE connectivities to H-22 as well as several of the protons in the prenyl moiety (H-18, H-19, H-19' and H-20), strongly indicating the positioning of the prenyl group and H-6 on the other side of the ring system as compared to H-3 and H-16.

**Table 3.** NMR spectroscopic data (500 and 800 MHz, DMSO-$d_6$) for acu-dioxomorpholine.

| Position | $\delta_H$ (*J* in Hz) | $\delta_C$ | HMBC | Noesy |
|---|---|---|---|---|
| 1 | - | 168.0 | - | - |
| 2 | 2.96 (3H, s) | 32.7 | 6, 8 | 6,9 |
| 3 | 5.32 (1H, dd, 8.8, 3.4) | 78.1 | 4, 22, 23 | 16, 22, 22', 24/28 |
| 4 | - | 163.5 | - | - |
| 5 | - | - | - | - |
| 6 | 5.43 (1H, s) | 81.9 | 2, 4, 8, 13, 14, 15, 16, 17 | 12, 18, 19, 19', 20, 22 |
| 7 | - | - | - | - |
| 8 | - | 150.8 | - | - |
| 9 | 6.44 (1H, d, 7.6) | 105.6 | 11, 13 | 2, 10 |
| 10 | 7.09 (1H, td, 7.6, 0.8) | 128.5 | 8, 12 | 9, 11 |
| 11 | 6.64 (1H, t, 7.3) | 116.9 | 9, 12, 13 | 10, 12 |
| 12 | 7.17 (1H, d, 7.3) | 124.1 | 8, 10, 14 | 11, 15, 18, 20, 21 |
| 13 | - | 128.8 | - | - |
| 14 | - | 59.6 | - | - |
| 15 | 2.43 (1H, dd, 12.7, 6.6) | 36.6 | 1, 6, 13, 14, 16 | 12, 15', 16 |
| 15' | 2.21 ( 1H, dd, 12.7, 11.1) | 36.6 | 1, 13, 14, 16 | 15, 16, 18, 20, 21 |
| 16 | 4.17 (1H, dd, 11.1, 6.6) | 56.7 | 1, 15 | 3, 15, 15' |
| 17 | - | 40.1 | - | - |
| 18 | 5.87 (1H, dd, 17.4, 11.0) | 143.3 | 14, 17, 21 | 6, 12, 15', 19, 19', 20, 21 |
| 19 | 5.02 (1H, dd, 17.4, 1.1) | 113.7 | 14, 17, 18 | 6, 18, 19', 20, 21 |
| 19' | 5.06 (1H, dd, 11.0, 1.1) | 113.6 | 14, 17, 18 | 6, 18, 19, 20, 21 |
| 20 | 0.82 (3H, s) | 22.4 | 14, 17, 18, 21 | 6, 12, 15', 18, 19, 19', 21 |
| 21 | 0.97 (3H, s) | 21.7 | 14, 17, 18, 20 | 12, 15', 18, 19, 19', 20 |
| 22 | 2.98 (1H, dd, 14.8, 8.8) | 34.9 | 3, 4, 23, 28 | 3, 6, 22', 24/28 |
| 22' | 3.33 (1H, m) | 34.9 | 3, 4, 23, 28 | 3, 22, 24/28 |
| 23 | - | 136.3 | - | - |
| 24 | 7.26 (1H, m) | 129.2 | 22, 26, 28 | 3, 22, 22' |
| 25 | 7.28 (1H, m) | 127.7 | - | 26 |
| 26 | 7.22 (1H, tt, 7.2, 1.6) | 126.2 | 28 | 25 |
| 27 | 7.28 (1H, m) | 127.7 | 23, 25 | - |
| 28 | 7.26 (1H, m) | 129.2 | - | 3, 22, 22' |

**Table 4.** NMR spectroscopic data (500 and 800 MHz, DMSO-$d_6$) for okaramine S.

| Position | $\delta_H$ (*J* in Hz) | $\delta_C$ | Hmbc | Noesy |
|---|---|---|---|---|
| 1 | 7.64 (1H, s) | 114.7 | 2, 4, 14, 15 | - |
| 2 | - | 123.8 | - | - |
| 4 | - | 137.6 | - | - |
| 5 | 4.26 (1H, q, 6.5) | 43.2 | 4, 14, 18 | 18 |
| 6 | - | 51.2 | - | - |
| 7 | 5.87 (1H, s) | - | 5, 9, 13, 17 | 9 |
| 8 | - | 144.5 | - | - |
| 9 | 6.90 (1H, d, 7.6) | 118.8 | 11, 13 | 10 |
| 10 | 6.99 (1H, t, 7.6) | 127.4 | 8, 12 | 9 |
| 11 | 6.65 (1H, t, 7.6) | 116.8 | 9, 13 | 12 |
| 12 | 7.74 (1H, d, 7.6) | 128.3 | 8, 10, 14 | 1,11 |
| 13 | - | 115.3 | - | - |
| 14 | - | 124.9 | - | - |
| 15 | - | 153.4 | - | - |
| 16 | 0.95 (3H, s) | 27.1 | 5, 6, 17 | 17 |
| 17 | 1.39 (3H, s) | 29.3 | 5, 6, 16 | 16 |
| 18 | 1.18 (3H, d, 7.2) | 18.1 | 4, 5, 6 | - |
| 19 | 7.76 (1H, s) | 117.5 | 20, 25, 32, 33, 34 | - |
| 20 | - | 125.2 | - | - |
| 22 | 5.90 (1H, d, 8.3) | 123.6 | 23, 24 | 23 |
| 23 | 6.15 (1H, d, 8.3) | 140.2 | 22, 24, 35/36 | 22 |
| 24 | - | 35.8 | - | - |
| 25 | - | 149.8 | - | - |
| 26 | 11.70 (1H, s) | - | 25, 32, 33 | - |
| 27 | - | 134.2 | - | - |
| 28 | 7.44 (1H, m) | 112.3 | - | 29/30 |
| 29 | 7.17 (1H, m) | 121.7 | 31 | 28, 31 |
| 30 | 7.17 (1H, m) | 121.7 | 31 | 28, 31 |
| 31 | 7.63 (1H, m) | 116.1 | 27, 29/30 | 12, 29/30 |
| 32 | - | 105.3 | - | - |
| 33 | - | 129.5 | - | - |
| 34 | - | 157.6 | - | - |
| 35 | 1.69 (3H, s) | 26.7 | 23, 24, 25, 35/36 | - |
| 36 | 1.69 (3H, s) | 26.7 | 23, 24, 25, 35/36 | - |

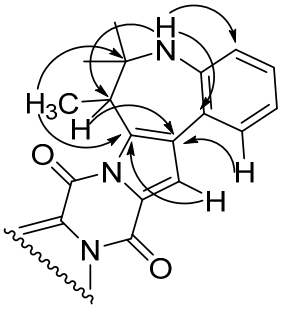
For okaramine S the molecular formula was by HRMS predicted to be $C_{32}H_{30}N_4O_2$. The azocinoindole part of the novel okaramine (Figure 7) is identical to that observed in okaramines A and H (Scheme 1), also produced by *A. aculeatus* [12]. The remaining part is highly similar, also originating from tryptophan and an isoprene unit, but in okaramine S, the terpene is fused with the ring system of the tryptophan moiety resulting in an intriguing, additional seven-membered ring, not previously reported in any okaramine. Key HMBCs from this part of okaramine S are depicted in Figure 8.

**Table 5.** NMR spectroscopic data (500 MHz, DMSO-$d_6$) for epi-10,23-dihydro-24,25-dehydroaflavinine.

| Position | $\delta_H$ (*J* in Hz) | $\delta_C$ | HMBC | H2BC | NOESY |
|---|---|---|---|---|---|
| 1 | 10.66 (s) | - | 2, 3, 4, 9 | 2 | 2,8 |
| 2 | 7.07 (d 2.0) | 122.7 | 3, 4, 9 | - | 1, 23, 25, 27 |
| 3 | - | 114.4 | - | - | |
| 4 | - | 126.7 | - | - | |
| 5 | 7.38 (d 7.8) | 116.8 | 7, 9 | 6 | 10, 11, 18 |
| 6 | 6.96 (t 7.2) | 117.7 | 4, 8 | 5, 7 | |
| 7 | 7.02 (t 7.3) | 120.0 | 5, 9 | 6, 8 | |
| 8 | 7.29 (d 8.0) | 110.9 | 4, 6 | 7 | 1 |
| 9 | - | 135.5 | - | - | |
| 10 | 3.59 (dd 13.3, 5.0) | 33.4 | 2, 3, 11, 12, 23, 24 | 11, 23 | 5, 11, 19, 25, 26 |
| 11 | 2.47 (m) | 37.6 | 10, 12, 23 | 10, 12 | 5, 10, 12, 13, 16, 19 |
| 12 | 1.27 (m) | 28.9 | - | 17 | 11 |
| 13 | 1.53 (m) | 28.1 | 29 | - | 11, 13', 16 |
| 13' | 0.81 (d 12.9) | 28.1 | - | - | 13 |
| 14 | 1.46 (m) | 27.4 | - | - | |
| 14' | 1.08 (d 13.7) | 27.4 | 27 | - | 28 |
| 15 | - | 38.2 | - | - | |
| 16 | 2.04 (m) | 30.4 | - | 17, 28 | 11, 13, 17', 28 |
| 17 | 1.71 (m) | 24.8 | - | 18 | |
| 17' | 1.22 (m) | 24.8 | - | - | |
| 18 | 1.99 (d 11.3) | 29.6 | - | - | 5, 19 |
| 18' | 1.74 (m) | 29.6 | - | - | 19, 30 |
| 19 | 4.64 (s) | 65.6 | - | - | 10, 11, 18, 18', 22, 26 |
| 20 | - | 42.9 | - | - | |
| 21 | 2.07 (m) | 23.5 | - | - | 29, 30 |
| 21' | 1.65 (m) | 23.5 | 19, 23 | - | 27 |
| 22 | 1.85 (m) | 26.6 | 20, 24 | 21, 23 | 19, 30 |
| 22' | 1.54 (m) | 26.6 | - | - | |
| 23 | 3.13 (m) | 42.3 | 24, 25, 26 | 10, 22 | 2, 25, 27 |
| 24 | - | 149.8 | - | - | |
| 25 | 4.81 (d 1.8) | 110.4 | 23, 26 | 26 | 2, 10, 23, 25' |
| 25' | 4.58 (d 1.8) | 110.4 | 23, 26 | 26 | 25, 26 |
| 26 | 1.45 (s) | 17.7 | 23, 24, 25 | - | 10, 19, 25' |
| 27 | 1.21 (s) | 21.1 | 11, 12 | 12 | 2, 21', 23 |
| 28 | 0.71 (d 6.7) | 15.4 | 15, 16, 17 | 16 | 14', 16, 17' |
| 29 | 0.92 (s) | 17.7 | 13, 15, 16, 20 | - | 17, 21, 30 |
| 30 | 4.28 (d 4.3) | - | 19, 20 | 19 | 18', 21, 22, 29 |

The molecular formula of the aflavinine analogue was predicted by HRMS to be $C_{28}H_{39}NO$ and dereplication indicated it could be aflavinine or an analogue. Since no aflavinine or analogue was previously described in *A. aculeatus*, this compound was also a target for isolation and structural elucidation. The NMR data, see Table 5, suggested the structure of 10,23-dihydro-24,25-dehydroaflavinine.

**Figure 8.** Key HMBCs substantiating the seven-membered ring in okaramine S.




10,23-Dihydro-24,25-dehydroaflavinine has earlier been reported from sclerotia from *A. flavus*, *A. parasiticus* and *A. tubingensis* [38]. The compound has not been reported from *A. aculeatus*, but it is known to be produced by the related black *Aspergillus A. costaricaensis* [39]. The stereochemistry of epi-10,23-dihydro-24,25-dehydroaflavinine was solved by the use of NOEs and *J*-couplings. The relative stereochemistry was found to be equal to that of 10,23-dihydro-24,25-dehydroaflavinine [38]. For investigation of the absolute configuration of the purified compound, optical rotation was measured. Two different values were obtained, $[\alpha] = 63.58°$ and $[\alpha] = 9.08°$ in MeOH and $CHCl_3$ respectively. These values were compared to the value of $[\alpha]_D = -1.20°$, which has earlier been reported for 10,23-dihydro-24,25-dehydroaflavinine in $CHCl_3$ [38]. Due to those opposite signs, our result suggests that the compound from *A. aculeatus* is the enantiomer of 10,23-dihydro-24,25-dehydroaflavinine, wherefore we have named it epi-10,23-dihydro-24,25-dehydroaflavinine. The 3D structure (see Supplementary Figure S40) was suggested by the isolated spin pair approximation [40].

### *2.3. Biological Testing of the Novel A. aculeatus Metabolites*

Aculenes A–C, acu-dioxomorpholine, okaramine S and epi-10,23-dihydro-24,25-dehydroaflavinine were tested for antifungal activity against *Candida albicans*. Endpoint optical density from compound screens were normalized with the negative controls and susceptibility evaluated as percentage reduction in optical density. None of the compounds showed significant antifungal activity.

### *2.4. Production of Sclerotia Reveals a Highly Regulated Metabolic Profile*

It has recently been demonstrated that sclerotium production can be prompted under specific conditions for a number of the black aspergilla [41]. This study showed that some metabolites are highly upregulated inside the sclerotia, compared to the metabolites in the mycelium. *A. aculeatus* is capable of producing sclerotia in vast amounts. Production was demonstrated on CYA and YES media, see Figure 9.

Analysis of extracts of sclerotia from MM, CYA and YES showed large upregulations of okaramines, see Figure 10. Aculenes A–C were still produced, though in smaller amounts. Furthermore the analysis showed that production of secalonic acids abolished on MM and CYA. The calbistrins were still produced and the same were acucalbistrins A–B, though in trace amounts. Some okaramine analogues have shown activity against silkworms [42–44] and the substantial upregulation of okaramines can be connected to sclerotium production as a defense mechanism to protect

against insects. Sclerotia are survival structures and the production is often provoked by extreme conditions [41,45].

**Figure 9.** The figure shows production of sclerotia on YES (left) and CYA media (right) after growth for 7 days at 25 °C in the dark. The sclerotia are the white to cream-colored structures seen among the black conidial *Aspergillus* heads.

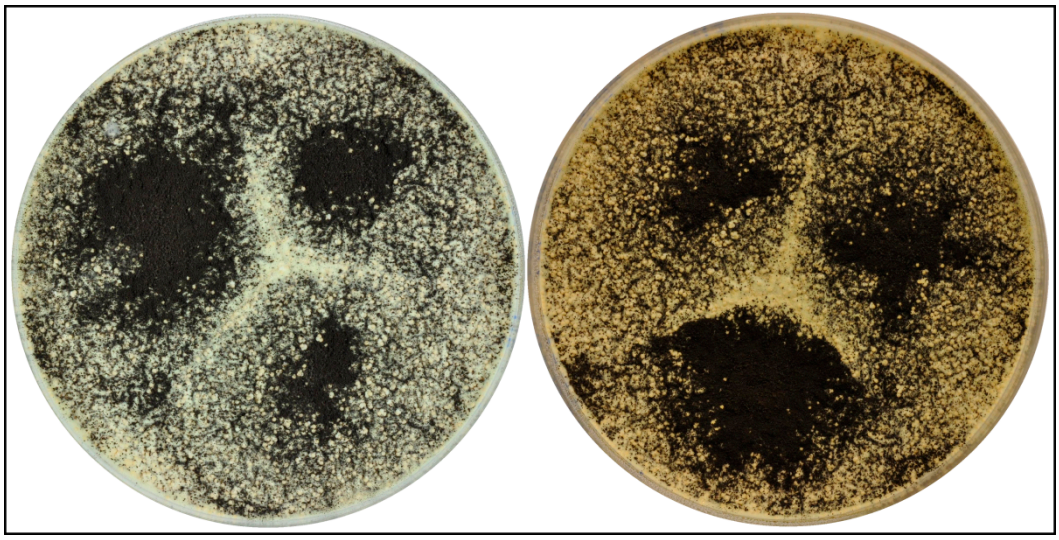
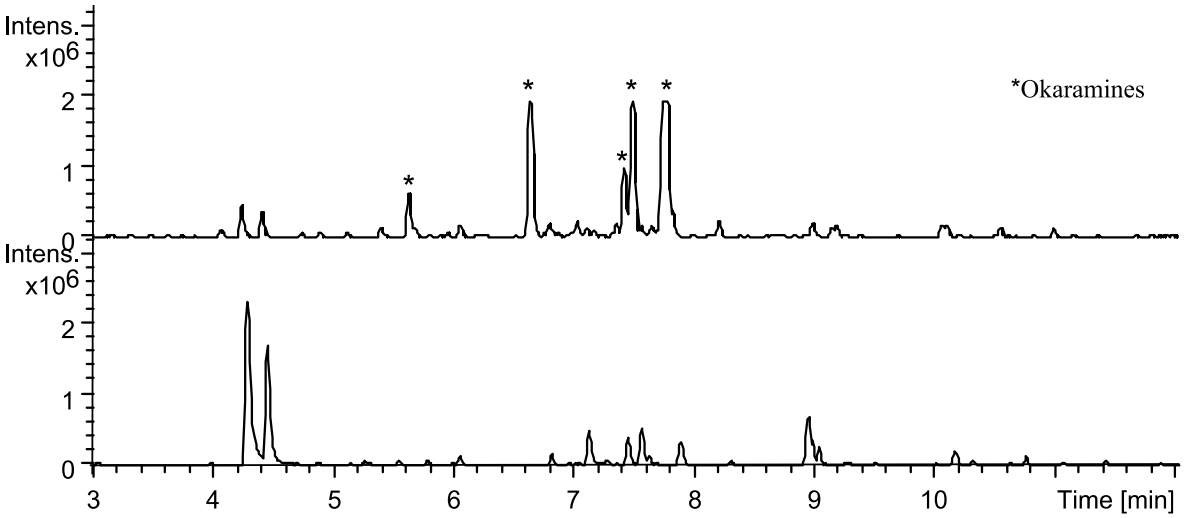


**Figure 10.** Base peak chromatograms of extraction from sclerotia (top) and plug extractions of *A. aculeatus* IBT 21030 after cultivation on YES agar for 7 days at 25 °C in the dark (bottom) for reference. The chromatograms are to scale. Stars indicate okaramines produced in the sclerotium extract.




## 3. Experimental Section

### *3.1. Fungal Growth, Strains and Media*

For standard screening the fungal strain was inoculated as three point stabs on solid media and incubated for seven days at 25 °C. Media was prepared as described by Samson *et al.* [25] The strain

IBT 21030 was compared to the culture ex type IBT 3244 = CBS 172.66 = ATCC 16872 = IMI 211388 = WB 5094, in order to authenticate the identity of the isolate.

### *3.2. Large Scale Extraction*

For the large scale preparation, *A. aculeatus* (IBT 21030) was inoculated as three point stabs on 200 plates of solid YES medium, and incubated in the dark for 7 days at 25 °C. The fungi were harvested and extracted. Ten Petri plates of fungi were transferred to separate stomacher bags. To each bag 100 mL of ethyl acetate (EtOAc) containing 1% formic acid (FA) were added and the bags were subsequently crushed using a stomacher for 30–60 s. After one hour the liquid phase was filtered and concentrated *in vacuo*. The stomacher bags still containing the fungi were again filled with EtOAc containing 1% FA, and then left overnight. The following day the solution was filtered and then concentrated *in vacuo*. The combined extract was dissolved in 800 mL methanol (MeOH) and $H_2O$ purified and deionized by a Millipore system through a 0.22 μm membrane filter (MQ $H_2O$) (9:1) and 800 mL of heptane added, whereafter the phases were separated. To the MeOH/MQ $H_2O$ phase 800 mL MQ $H_2O$ was added, and metabolites were then extracted with 4 × 400 mL dichloromethane (DCM). The phases were then concentrated separately *in vacuo.*

### *3.3. Plug Extraction*

For standard screening three plugs were taken from one colony by use of a 6 mm plug drill; one from the center of the colony, one from the edge near the other colonies and one from the edge as far away from the other colonies. The plugs were transferred to 2 mL vials and 500 μL extraction solvent were added. A mixture of MeOH/DCM/EtOAc (1:2:3 v/v/v) containing 1% FA was used. To each sample 40 μL chloramphenicol in ethanol (500 μg/mL) was added as an internal standard. The vials were placed in an ultrasonic bath (Branson 2510 or 3520) for 60 min. The extract was then transferred to a clean vial and evaporated to dryness. This was either achieved by leaving the vials in a fume hood over night or by applying nitrogen airflow at 25–32 °C. After evaporation 500 μL MeOH was added and the sample was then ultrasonicated for 20 minutes. The extract was then filtered using a 0.45 μm PTFE filter.

### *3.4. Sclerotium Extraction*

*A. aculeatus* IBT 21030 was three-point inoculated on YES, CYA and MM plates and incubated at 30 °C for ten days and then at room temperature for eight days. Sclerotia were harvested by applying MQ $H_2O$ to the plate and carefully harvest with Drigalski spatula. The liquid including the sclerotia and spores was filtred through sterilised Miracloth and washed with MQ $H_2O$, allowing for the spores to pass through, while sclerotia were staying in the filter. Sclerotia were transferred to true Eppendorf tubes and washed with water to remove remaining spores. After evaporation of excess MQ $H_2O$, two big and two small steel balls was added to each tube and the tubes were shaken at 2000 rpm for 2 × 60 s. To each tube 40 μL chloramphenicol (internal standard) and 1 mL MeOH/DCM/EtOAc (1:2:3 v/v/v) containing 1% FA was added and the extract was transferred to clean 2 mL vials. The samples were ultrasonicated for 1 h and the extract was transferred to clean vials and solvent was evaporated by applying airflow at

25–32 °C. The samples were redissolved in 500 µL MeOH, ultrasonicated for 20 min and filtrated to fresh vials using 0.45 µm PTFE filters.

### *3.5. UHPLC-DAD-HRMS Analysis*

Analysis was performed using ultra-high-performance liquid chromatography (UPHLC) UV/Vis diode array detector (DAD) high-resolution MS (TOFMS) on a maXis 3G orthogonal acceleration quadrupole time of flight mass spectrometer (Bruker Daltonics, Bremen, Germany) equipped with an electrospray ionization (ESI) source and connected to an Ultimate 3000 UHPLC system (Dionex, Sunnyvale, CA, USA). The column used was a reverse-phase Kinetex 2.6-µm $C_{18}$, 100 × 2.1 mm (Phenomenex, Torrance, CA, USA), and the column temperature was maintained at 40 °C. A linear water-acetonitrile (ACN) (LCMS-grade) gradient was used (both solvents were buffered with 20 mM FA) starting from 10% (vol/vol) ACN and increased to 100% in 10 min, maintaining this rate for 3 min before returning to the starting conditions in 0.1 min and staying there for 2.4 min before the following run. A flow rate of 0.4 $mL \cdot min^{-1}$ was used. TOFMS was performed in ESI+ with a data acquisition range of 10 scans per second at *m/z* 100–1000. The TOFMS was calibrated using Bruker Daltonics high precision calibration algorithm (HPC) by means of the use of the internal standard sodium formate, which was automatically infused before each run. UV/VIS spectra were collected at wavelengths from 200 to 700 nm. Data processing was performed using DataAnalysis 4.0 software (Bruker Daltonics).

### *3.6. NMR*

The 1D and 2D spectra were recorded on a Varian Unity Inova-500 MHz spectrometer located at DTU, or on a Bruker Avance 800 MHz spectrometer located at the Danish Instrument Centre for NMR Spectroscopy of Biological Macromolecules at Carlsberg Laboratory. Spectra were acquired using standard pulse sequences. The deuterated solvent was DMSO-$d_6$ and signals were referenced by solvent signals for DMSO-$d_6$ at $\delta_H$ = 2.49 ppm and $\delta_C$ = 39.5 ppm. The NMR data was processed in MestReNova V.6.0.1–5391 or Bruker Topspin. Chemical shifts are in ppm ($\delta$) and scalar couplings are reported in hertz (Hz). The sizes of the *J* coupling constants reported in the tables are the experimentally measured values from the 1D $^1$H and DQF-COSY spectra. There are minor variations in the measurements which may be explained by the uncertainty of *J* and the spectral digital resolution.

Distances for the epi-10,23-dihydro-24,25-dehydroaflavinine were obtained from 2D NOESY experiments using the isolated spin pair approximation (ISPA) [40]. The linear range was increased by the method suggested by Macura *et al.* [46,47]. The used mixing time was 150 ms. Different mixing times were used to construct a buildup curve to ensure that only crosspeaks which fitted the ISPA were used.

The $^3J$-couplings from angles in the 3D structures of epi-10,23-dihydro-24,25-dehydroaflavinine and the aculenes were calculated by the Haasnoot-DeLeeuw-Altona (HLA) equation and by DFT computations of the final structure [48,49].

The simulations were conducted using the program Maestro (Version 9.3.515, MMshare Version 2.1.515) from the Schrödinger suite. Conformational searches in implicit solvents (DMSO) were run by MacroModel (version 9.9, Schrödinger, LLC, New York, NY, USA, 2012) using the force

fields OPLS2005 and MMFFs. Monte Carlo torsional sampling was used to generate the structures and the minimization method was PRCG. The number of steps were 20,000, and only conformations within 30 kJ/mol of the found minimum were considered. The solvent DMSO was treated as a constant dielectric constant of 47.0. Both force fields gave similar results.

Selected structures were further optimized by HF/3-21G using Jaguar (Jaguar, version 7.9, 2012, Schrödinger, LLC, New York, NY, USA,) followed by DFT (for epi-10,23-dihydro-24,25-dehydroaflavinine) [50]. DFT calculations were carried out by Gaussian09 [51] using cam-B3LYP/6-311++G(d,p) for optimization and B3LYP/6-31G(d,p) u+1s for calculating *J*-coupling constants (only Fermi Contact contribution considered, see reference [49]).

### *3.7. Marfey's Reaction*

Aculenes A and B (100 μg) were hydrolyzed with 6 M HCl (200 μL) in 2 mL analytical vials with lids, and were left at 110 °C for 22 h. The solvent was removed at a speedvac. For reference, 50 μL (2.5 μmol) standard L- and D-proline were prepared. To all vials 50 μL $H_2O$, 20 μL 1 M $NaHCO_3$ and 100 μL 1% 1-fluoro-2,4-dinitrophenyl-5-L-alanineamide (FDAA) in acetone were added. The vials were left for 40 °C for 1 h. 10 μL 2 M HCl was added followed by 820 μL MeOH and HPLC-DAD-MS data were acquired.

### *3.8. Purification of Metabolites*

The DCM phase from the large scale extraction consisting of 2.375 g was absorbed onto Sepra ZT C18 (Phenomenex) and dried before packing into a 50 g (~66 mL) SNAP column (Biotage, Uppsala, Sweden) with Septra ZT C18 material. The extract was then fractionated using an Isolera flash purification system (Biotage). The gradient started with 15:100 MeOH/$H_2O$ in the first three column volumes (CV), then went to 100% MeOH over 18 CVs and stayed here for 3 CVs using a flow rate of 40 mL/min. MeOH was of HPLC grade and $H_2O$ was purified and deionized by Millipore system through 0.22 μm membrane filter (MQ $H_2O$) and both were added 50 ppm trifluoroacetic acid (TFA). Fractions were automatically collected 1 CV at a time. The fractions were subjected to further purification on a semi-preparative HPLC, which was either a Waters 600 Controller with a 996 photodiode array detector (Waters, Milford, MA, USA) or a Gilson 322 Controller connected to a 215 Liquid Handler, 819 Injection Module and a 172 DAD (Gilson, Middleton, WI, USA). This was achieved using a Luna II C18 column (250 × 10 mm, 5 μm, Phenomenex). 50 ppm TFA was added to ACN of HPLC grade and MQ $H_2O$. For choice of system, flow rate, gradients and yields see descriptions for the specific compound.

*Aculenes A and B.* The three first Isolera flash chromatography fractions, eluding with 10%–25% MeOH, were subjected to further purification on the waters semi-preparative HPLC. A linear water-ACN gradient was used starting with 15% ACN and increasing to 100% over twenty minutes using a flow rate of 4 mL/min. Two compounds eluting with 54% and 56% ACN respectively, were collected. This yielded two pure compounds, 3.0 mg of aculene A and 3.7 mg of aculene B.

*Aculene A*: HRMS: *m/z* = 316.1910. $[M+H]^+$, calculated for $[C_{19}H_{25}NO_3+H]^+$: *m/z* = 316.1907. $[\alpha]_D^{20}$ = +0.63° (MeOH). $^1$H- and $^{13}$C-NMR (see Tables 1 and 2).

*Aculene B*: HRMS: *m/z* = 318.2068. $[M+H]^+$, calculated for $[C_{19}H_{27}NO_3+H]^+$: *m/z* = 318.2064. $[\alpha]_D^{20}$ = +6.96° (MeOH). $^1$H- and $^{13}$C-NMR (see Tables 1 and 2).

*Aculene C*. One of the Isolera flash chromatography fractions, eluding with 50% MeOH, was subjected to further purification on the waters semi-preparative HPLC. An isocratic water-ACN gradient was used starting with 40% ACN over twenty minutes using a flow rate of 4 mL/min. A compound eluding with 40% ACN was collected. This yielded 2.0 mg of aculene C.

HRMS: *m/z* = 319.1382 $[M+H]^+$, calculated for $[C_{14}H_{18}O_2+H]^+$: *m/z* = 319.1379. $[\alpha]_D^{20}$ = 0.00° (MeOH). $^1$H- and $^{13}$C-NMR (see Tables 1 and 2).

*Calbistrin A and C*. One of the Isolera flash chromatography fractions, eluding with 100% MeOH, was subjected to further purification on the waters semi-preparative HPLC. An isocratic water-ACN method was used with 52% ACN over eighteen minutes using a flow rate of 4 mL/min. Two compounds eluding at 52% ACN were collected. This yielded 5.2 mg of calbistrin A and 4.6 mg of calbistrin C.

*Calbistrin A*: HRMS: *m/z* = 563.2622 $[M+Na]^+$, calculated for $[C_{31}H_{40}O_8+Na]^+$: *m/z* = 563.2615. $[\alpha]_D^{20}$ = +46.1° (MeOH). $^1$H-NMR (499.87 MHz, DMSO-*d*$_6$, 25 °C, 2.49 ppm): 0.77 (3H, d, *J* = 7.1 Hz), 0.97 (3H, d, *J* = 7.1 Hz), 1.13 (3H, s), 1.22 (3H, s), 1.32 (1H, m), 1.69 (3H, s), 1.98 (3H, s), 2.02 (1H, m), 2.30 (1H, dd, *J* = 13.8, 3.8 Hz), 2.37 (1H, m), 2.41 (1H, m), 2.69 (1H, dd, *J* = 13.8, 8.2 Hz), 2.80 (1H, m), 3.94 (1H, d, *J* = 9.4 Hz), 5.06 (1H, m, -OH), 5.10 (1H, m), 5.59 (1H, d, *J* = 9.6 Hz), 5.70 (1H, s), 5.86 (1H, m), 5.88 (1H, d, *J* = 14.9 Hz), 5.97 (1H, d, *J* = 9.8 Hz), 6.05 (1H, d, *J* = 11.2 Hz), 6.30 (1H, d, *J* = 12.0 Hz), 6.38 (1H, d, *J* = 15.1 Hz), 6.68 (1H, dd, *J* = 15.1, 11.2 Hz), 7.51 (1H, dd, *J* = 14.9, 12.0 Hz; $^{13}$C-NMR (125.70 MHz, DMSO-*d*$_6$, 25 °C, 39.5 ppm): 10.9, 12.3, 13.0, 13.7, 20.6, 25.4, 26.0, 34.7, 39.8, 43.6, 43.7, 56.2, 56.3, 68.3, 78.3, 121.2, 126.4, 127.0, 127.2, 130.8, 132.7, 133.8, 135.7, 139.4, 140.9, 143.4, 173.9, 212.8.

*Calbistrin C:* HRMS: *m/z* = 565.2776 $[M+Na]^+$, calculated for $[C_{31}H_{42}O_8+Na]^+$: *m/z* = 565.2772. $[\alpha]_D^{20}$ = +8.67° (MeOH); $^1$H-NMR (499.87 MHz, DMSO-*d*$_6$, 25 °C, 2.49 ppm): 0.77 (3H, d, *J* = 7.1 Hz), 0.95 (3H, d, *J* = 7.1 Hz), 1.01 (3H, s), 1.21 (1H, m), 1.27 (3H, s), 1.71 (3H, s), 1.93 (1H, m), 1.98 (3H, s), 2.35 (1H, m), 2.38 (1H, qin, *J* = 8.4), 2.56 (1H, m), 2.92 (1H, dt, *J* = 17.5, 7.5 Hz), 2.99 (1H, m), 3.56 (2H, m), 3.95 (1H, dd, *J* = 9.4, 3.1 Hz), 5.09 (1H, br.d, *J* = 3.5 Hz, -OH), 5.05 (1H, s, -OH), 5.20 (1H, m), 5.37 (1H, d, *J* = 9.9 Hz), 5.59 (1H, m), 5.88 (1H, d, *J* = 14.9 Hz), 5.90 (1H, d, *J* = 9.9 Hz), 6.07 (1H, d, *J* = 11.1 Hz), 6.31 (1H, d, *J* = 12.0 Hz), 6.39 (1H, d, *J* = 15.1 Hz), 6.69 (1H, dd, *J* = 15.1, 11.1 Hz), 7.51 (1H, dd, *J* = 14.9, 12.0); $^{13}$C-NMR (125.70 MHz, DMSO-*d*$_6$, 25 °C, 39.5 ppm): 10.9, 12.3, 13.0, 13.7, 20.6, 25.4, 26.0, 34.7, 39.8, 43.6, 43.7, 56.2, 56.3, 68.3, 78.3, 121.2, 126.4, 127.0, 127.2, 130.8, 132.7, 133.8, 135.7, 139.4, 140.9, 143.4, 173.9, 212.8.

*Acu-dioxomorpholine*. One of the Isolera flash chromatography fractions, eluding with 95% ACN, was subjected to further purification on the waters semi-preparative HPLC). A linear water-ACN gradient was used starting with 60% ACN increasing to 100% over twenty minutes using a flow rate of 4 mL/min. A compound eluding with 85% ACN was collected. This yielded 2.1 mg of acu-dioxomorpholine. HRMS: *m/z* = 417.2176 $[M+H]^+$, calculated for $[C_{26}H_{29}N_2O_3+H]^+$: *m/z* = 417.2173. $[\alpha]_D^{20}$ = −49.23° (MeOH). $^1$H- and $^{13}$C-NMR (see Table 3).

*Epi-10,23-dihydro-24,25-dehydroaflavinine*. One Isolera flash chromatography fraction eluding with 100% MeOH, was subjected to purification on the waters semi-preparative HPLC. A gradient of 15%–100% ACN over twenty minutes was used and the flow rate was 4 mL/min. A compound eluding with 100% ACN was collected. This yielded 8.7 mg of epi-10,23-dihydro-24,25-dehydroaflavinine. HRMS: *m/z* = 406.3113 $[M+H]^+$, calculated for $[C_{28}H_{39}NO+H]^+$: *m/z* = 406.3104. $[\alpha]_D^{20}$ = +63.58° (MeOH); $^1$H- and $^{13}$C-NMR (see Table 5).

*Okaramine S*. One Isolera flash chromatography fraction, eluding with 100% MeOH, was subjected to purification on the Gilson semi-preparative HPLC. This was done using a gradient of 70%–100% ACN over fifteen minutes using a flow rate was 5 mL/min. A compound eluding with 80% ACN was collected. This yielded 1.7 mg of the pure compound. HRMS: *m/z* = 503.2447 $[M+H]^+$, calculated for $[C_{32}H_{30}N_4O_2+H]^+$: *m/z* = 503.2441. $[\alpha]_D^{20}$ = −15.29° (MeOH); $^1$H- and $^{13}$C-NMR (see Table 4).

*Okaramine J*. One Isolera flash chromatography fraction, eluding with 74% MeOH, was subjected to purification on the waters semi-preparative HPLC. This was done using a gradient of 50%–100% ACN over twenty minutes using a flow rate of 4 mL/min. A compound eluding with 82% ACN was collected. This yielded 2.5 mg of the pure compound. HRESIMS: *m/z* = 525.2865 $[M+H]^+$, calculated for $[C_{32}H_{36}N_4O_3+H]^+$: *m/z* = 525.2860. $[\alpha]_D^{20}$ = +15.38° (MeOH); $^1$H-NMR (499.87 MHz, DMSO-$d_6$, 25 °C, 2.49 ppm): 1.52 (6H, s, H-17, H-18), 1.68 (3H, s, H-34), 1.73 (3H, s, H-35), 1.86 (1H, m, H-20), 2.42 (1H, dd, 13.2, 6.8, H-20'), 2.99 (1H, dd, 15.2, 9.4, H-1), 3.14 (1H, dd, 16.3, 7.1, H-31), 3.23 (1H, dd, 16.8, 7.5, H-31'), 3.59 (1H, dd, 15.2, 4.2, H-1'), 4.46 (1H, dd, 9.1, 4.4, H-2), 4.67 (1H, dd, 11.3, 6.4, H-19), 5.05 (1H, dd, 10.5, 1.1, H-4), 5.08 (1H, dd, 17.4, 1.1, H-4'), 5.26 (1H, t, 7.4, H-32), 5.33 (1H, d, 4.5, H-29), 6.01 (1H, s , H-36), 6.12 (1H, d, 4.4, H-28), 6.22 (1H, dd, 17.4, 10.5, H-5), 6.32 (1H, s, H-3), 6.67 (1H, t, 7.5, H-24), 6.88 (1H, d, 7.5, H-25), 6.96 (1H, t, 7.5, H-12), 6.96 (H-26), 7.05 (2H, H-11, H-23), 7.35 (1H, s, H-10), 7.53 (1H, s, H-13), 10.67 (1H, s, H-8); $^{13}$C-NMR (125.70 MHz, DMSO-$d_6$, 25 °C, 39.5 ppm): 17.5 (C-34), 24.7 (C-1), 25.2 (C-35), 27.6 (C-17,C-18), 28.2 (C-31), 39.0 (C-6), 40.8 (C-20), 55.1 (C-2), 58.2 (C-19), 83.5 (C-29), 85.5 (C-21), 104.3 (C-15), 110.7 (C-10), 110.9 (C-4), 117.6 (C-13), 118.4 (C-12, C-24), 118.9 (C-23), 120.0 (C-11), 121.5 (C-32), 122.7 (C-22), 128.3 (C-14), 131.0 (C-26), 132.0 (C-33), 134.5 (C-9), 141.1 (C-7), 145.8 (C-27), 146.0 (C-5), 167.5 (C-16), 169.7 (C-30).

### *3.9. Antifungal Susceptibility Testing*

All compounds were screened for antifungal activity towards *Candida albicans* in accordance with the CLSI standards using RPMI-1640 medium adjusted to pH 7 with 0.165 M MOPS buffer [52]. The screening was performed as described by Holm *et al.* [53].

## 4. Conclusions

Investigation of the chemical profile of *Aspergillus aculeatus* by UHPLC-DAD-HRMS has identified several novel compounds some of which have been selected, purified and their structure elucidated by 1D and 2D NMR spectroscopy. A novel class of related compounds – the aculenes – has been discovered. Aculenes A and B are hybrids containing the amino acid proline but also a fourteen carbon moeity we hypothesize to originate from a sesquiterpene with the loss of one carbon atom. Aculene C is believed to be a precursor to aculene A, missing the proline part of the molecule.

Moreover there is UHPLC-DAD-HRMS evidence suggesting a fourth structure, aculene D, a possible precursor to aculene B. Two other larger, however chemically unstable metabolites have been discovered. As they apparently were degrading to two known parts, aculene A and B respectively and calbistrin A, a tentative suggestion for their structures have been given. Two further novel compounds, acu-dioxomorpholine and okaramine S has also been discovered to be produced by *A. aculeatus*. Acu-dioxomorpholine has a remarkable structure, reminding of diketopiperazine, but with a lactone instead of the one lactame part and okaramine S displays a seven-membered ring, not previously seen in any reported okaramine. Furthermore the aflavinine analogue epi-10,23-dihydro-25,25-dehydroaflavinine was detected. Production of sclerotia was observed under specific conditions, and here an upregulation of okaramines was observed. Altogether the chemical profile of *Aspergillus aculeatus* has been examined and several novel compounds have been characterized.

## Supplementary Materials

$^{1}$H, DQF-COSY, HSQC, HMBC, NOESY and UV spectra for the novel compounds as well as suggestion for the structures of acucalbistrin A and B and elucidation of the stereochemistry of aculenes A–D and epi-10,23-dihydro-24,25-dehydroaflavinine are available in the supplementary. Supplementary materials can be accessed at: http://www.mdpi.com/1420-3049/19/8/10898/s1.

## Acknowledgments

We thank the Danish Instrument Center for NMR Spectroscopy of Biological Macromolecules for 800 MHz NMR time. This work was supported by grant 09-064967 from the Danish Council for Independent Research, Technology, and Production Sciences. We thank James B. Gloer for supplying us with an authentic standard of 10,23-dihydro-24,25-dehydroaflavinine. We thank Peter B. Knudsen for running the antifungal assay.

## Author Contributions

LMP executed the experiments, isolated and elucidated the structures of the compounds and wrote the paper. CH assisted in purification and structural elucidation and performed the investigation of the stereochemistry. JCF assisted in the sclerotial study. CHG assisted with structural elucidations. TOL planned the study and guided preparation of the manuscript.

## Conflicts of Interest

The authors declare no conflict of interest.

## References

1. Nielsen, K.F.; Mogensen, J.M.; Johansen, M.; Larsen, T.O.; Frisvad, J.C. Review of secondary metabolites and mycotoxins from the *Aspergillus niger* group. *Anal. Bioanal. Chem.* **2009**, *395*, 1225–1242.

2. Song, Y.C.; Li, H.; Ye, Y.H.; Shan, C.Y.; Yang, Y.M.; Tan, R.X. Endophytic naphthopyrone metabolites are co-inhibitors of xanthine oxidase, SW1116 cell and some microbial growths. *FEMS Microbiol. Lett.* **2004**, *241*, 67–72.
3. Koyama, K.; Ominato, K.; Natori, S.; Tashiro, T.; Tsuruo, T. Cytotoxicity and antitumor activities of fungal bis(naphtho-gamma-pyrone) derivatives. *J. Pharmacobio-Dyn.* **1988**, *11*, 630–635.
4. Abarca, M.; Bragulat, M. Ochratoxin A production by strains of Aspergillus niger var. niger. *Appl. Environ. Microbiol.* **1994**, *60*, 2650–2652.
5. Frisvad, J.C.; Smedsgaard, J.; Samson, R.A.; Larsen, T.O.; Thrane, U. Fumonisin B2 Production by *Aspergillus niger*. *J. Agric. Food Chem.* **2007**, *55*, 9727–9732.
6. Varga, J.; Frisvad, J.C.; Kocsubé, S.; Brankovics, B.; Tóth, B.; Szigeti, G.; Samson, R.A. New and revisited species in *Aspergillus* section *Nigri*. *Stud. Mycol.* **2011**, *69*, 1–17.
7. Jurjević, Z.; Peterson, S.W.; Stea, G.; Solfrizzo, M.; Varga, J.; Hubka, V.; Perrone, G. Two novel species of *Aspergillus* section *Nigri* from indoor air. *IMA Fungus* **2012**, *3*, 159–173.
8. Mizuno, K.; Yagi, A.; Satoi, S.; Takada, M.; Hayashi, M.; Asano, K.; Matsuda, T. Studies on aculeacin. I. Isolation and characterization of aculeacin A. *J. Antibiot.* **1977**, *4*, 297–302.
9. Satoi, S.; Yagi, A.; Asano, K.; Misuno, K.; Watanabe, T. Studies on aculeacin. II. Isolation and characterization of aculeacins B, C, D, E, F and G. *J. Antibiot.* **1977**, *30*, 303–307.
10. Watanabe, S.; Hirai, H.; Ishiguro, M.; Kambara, T.; Kojima, Y.; Matsunaga, T.; Nishida, H.; Suzuki, Y.; Harwood, H.J.; Huang, L.H.; *et al.* CJ-15,183, a new inhibitor of squalene synthase produced by a fungus, *Aspergillus aculeatus*. *J. Antibiot.* **2001**, *54*, 904–910.
11. Andersen, R.; Buechi, G.; Kobbe, B.; Demain, A.L. Secalonic acids D and F are toxic metabolites of *Aspergillus aculeatus* . *J. Org. Chem.* **1977**, *42*, 352–353.
12. Hayashi, H.; Furutsuka, K.; Shiono, Y. Okaramines H and I, new Okaramine Congeners, from *Aspergillus aculeatus* . *J. Nat. Prod.* **1999**, *62*, 315–317.
13. Ingavat, N.; Mahidol, C.; Ruchirawat, S.; Kittakoop, P. Asperaculin A, a sesquiterpenoid from a marine-derived fungus, Aspergillus aculeatus. *J. Nat. Prod.* **2011**, *74*, 1650–1652.
14. Ingavat, N.; Dobereiner, J.; Wiyakrutta, S.; Mahidol, C.; Ruchirawat, S.; Kittakoop, P. Aspergillusol A, an alpha-glucosidase inhibitor from the marine-derived fungus Aspergillus aculeatus. *J. Nat. Prod.* **2009**, *72*, 2049–2052.
15. Chen, L.; Zhang, Q.-Q.; Zhang, W.-W.; Liu, Q.-Y.; Zheng, Q.-H.; Zhong, P.; Hu, X.; Fang, Z.-X. Aculeatusquinones A-D, novel metabolites from the marine-derived fungus Aspergillus aculeatus. *Heterocycles* **2013**, *87*, 861–868.
16. Goldberg, I.; Rokem, J.; Pines, O. Organic acids: Old metabolites, new themes. *J. Chem. Technol. Biotechnol.* **2006**, *81*, 1601–1611.
17. Pel, H.J.; de Winde, J.H.; Archer, D.B.; Dyer, P.S.; Hofmann, G.; Schaap, P.J.; Turner, G.; de Vries, R.P.; Albang, R.; Albermann, K.; *et al.* Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88. *Nat. Biotechnol.* **2007**, *25*, 221–231.
18. Perrone, G.; Susca, A.; Cozzi, G.; Ehrlich, K.C.; Varga, J.; Frisvad, J.C.; Meijer, M.; Noonim, P.; Mahakarnchanakul, W.; Samson, R.A. Biodiversity of Aspergillus species in some important agricultural products. *Stud. Mycol.* **2007**, *59*, 53–66.

19. Adisa, V.A. Pectic enzymes of Aspergillus aculeatus associated with post-harvest deterioration of citrus sinensis fruit. *J. Food Biochem.* **1989**, *13*, 243–252.
20. Bhat, M.K. Cellulases and related enzymes in biotechnology. *Biotechnol. Adv.* **2000**, *18*, 355–383.
21. Fujimoto, H.; Ooi, T.; Wang, S.; Takizawa, T. Purification and properties of three xylanases from Aspergillus aculeatus. *Biosci. Biotechnol. Biochem.* **1995**, *59*, 538–540.
22. Polizeli, M.L.T.M.; Rizzatti, A.C.S.; Monti, R.; Terenzi, H.F.; Jorge, J.A.; Amorim, D.S. Xylanases from fungi: Properties and industrial applications. *Appl. Microbiol. Biotechnol.* **2005**, *67*, 577–591.
23. Olutiola, P.O.; Nwaogwugwu, R.I. Growth, sporulation and production of maltase and proteolytic enzymes in Aspergillus aculeatus. *Trans. Br. Mycol. Soc.* **1982**, *78*, 105–113.
24. Christophersen, C.; Crescente, O.; Frisvad, J.C.; Gram, L.; Nielsen, J.; Nielsen, P.H.; Rahbaek, L. Antibacterial activity of marine-derived fungi. *Mycopathologia* **1999**, *143*, 135–138.
25. Samson, R.A.; Houbraken, J.; Thrane, U.; Frisvad, J.C.; Andersen, B. *Food and Indoor Fungi*; CBS-KNAW Fungal Biodiversity Centre: Utrecht, The Netherlands, 2010.
26. Smedsgaard, J. Micro-scale extraction procedure for standardized screening of fungal metabolite production in cultures. *J. Chromatogr. A* **1997**, *760*, 264–270.
27. Nielsen, K.F.; Ma, M.; Rank, C.; Frisvad, J.C.; Larsen, T.O. Dereplication of microbial natural products by LC-DAD-TOFMS. *J. Nat. Prod.* **2011**, *74*, 2338–2348.
28. Laatsch, H. Antibase 2012. Available online: http://www.wiley-vch.de/stmdata/antibase.php (accessed on 1 February 2014).
29. Shiono, Y.; Akiyama, K.; Hayashi, H. New okaramine congeners, okaramines J, K, L, M and related compounds, from *Penicillium Simplicissimum* ATCC90288. *Biosci. Biotechnol. Biochem.* **1999**, *63*, 1910–1920.
30. Nielsen, K.F.; Smedsgaard, J. Fungal metabolite screening: Database of 474 mycotoxins and fungal metabolites for dereplication by standardised liquid chromatography-UV-mass spectrometry methodology. *J. Chromatogr. A* **2003**, *1002*, 111–136.
31. Jackson, M.; Karwowski, J.; Humphrey, P. Calbistrins, novel antifungal agents produced by Penicillium restrictum. I: Production, taxonomy of the producing organism and biological activity. *J. Antibiot.* **1993**, *46*, 34–38.
32. Brill, G.; Chen, R.; Rasmussen, R. Calbistrins, novel antifungal agents produced by Penicillium restrictum. II: Isolation and elucidation of structure. *J. Antibiot.* **1993**, *46*, 39–47.
33. Sørensen, A.; Lübeck, P.S.; Lübeck, M.; Nielsen, K.F.; Ahring, B.K.; Teller, P.J.; Frisvad, J.C. Aspergillus saccharolyticus sp. nov., a black Aspergillus species isolated in Denmark. *Int. J. Syst. Evol. Microbiol.* **2011**, *61*, 3077–3083.
34. Marfey, P. Determination of D-amino acids. II. Use of a bifunctional reagent, 1,5-di-fluoro-2,4-dinitrobenzen. *Carlsb. Res. Commun.* **1984**, *49*, 591–596.
35. Rank, C.; Klejnstrup, M.L.; Petersen, L.M.; Kildgaard, S.; Frisvad, J.C.; Gotfredsen, C.H.; Larsen, T.O. Comparative chemistry of Aspergillus oryzae (RIB40) and A. flavus (NRRL 3357). *Metabolites* **2012**, *2*, 39–56.

36. Boyes-Korkis, J.; Gurney, K.; Penn, J. Anacine, a new benzodiazepine metabolite of Penicillium aurantiogriseum produced with other alkaloids in submerged fermentation. *J. Nat. Prod.* **1993**, *56*, 1707–1717.
37. Wang, H.J.; Gloer, J.B.; Wicklow, D.T.; Dowd, P.F. Mollenines A and B: new dioxomorpholines from the ascostromata of Eupenicillium molle. *J. Nat. Prod.* **1998**, *61*, 804–807.
38. TePaske, M.R.; Gloer, J.B.; Wicklow, D.T.; Dowd, P.F. Three new aflavinines from the sclerotia of *Aspergillus tubingensis*. *Tetrahedron* **1989**, *45*, 4961–4968.
39. Samson, R.A.; Noonim, P.; Meijer, M.; Houbraken, J.; Frisvad, J.C.; Varga, J. Diagnostic tools to identify black aspergilli. *Stud. Mycol.* **2007**, *59*, 129–145.
40. Butts, C.P.; Jones, C.R.; Towers, E.C.; Flynn, J.L.; Appleby, L.; Barron, N.J. Interproton distance determinations by NOE--surprising accuracy and precision in a rigid organic molecule. *Org. Biomol. Chem.* **2011**, *9*, 177–184.
41. Frisvad, J.C.; Petersen, L.M.; Lyhne, E.K.; Larsen, T.O. Formation of sclerotia and production of indoloterpenes by *Aspergillus niger* and other species in section *Nigri*. *PLoS One* **2014**, *9*, e94857.
42. Murao, S.; Hayashi, H.; Takiuchi, K.; Arai, M. Okaramine A, a novel indole alkaloid with insecticidal activity, from Penicillium simplicissimum AK-40. *Agric. Biol. Chem.* **1988**, *52*, 885–886.
43. Hayashi, H.; Takiuchi, K.; Murao, S.; Arai, M. Okaramine B, an insecticidal indole alkaloid, produced by Penicillium simplicissimum AK-40. *Agric. Biol. Chem.* **1988**, *52*, 2131–2133.
44. Hayashi, H.; Asabu, Y.; Murao, S.; Arai, M. New Okaramine Congeners, Okaramins D, E, and F, from Penicillium simplicissimum ATCC 90288. *Jpn. Soc. Biosci. Biotechnol. Agrochem.* **1995**, *59*, 246–250.
45. Wicklow, D.T. Metabolites in the coevolution of fungal chemical defence systems. In *Coevolution of Fungi with Plant and Animals*; Pirozynski, K., Hawksworth, D., Eds.; Academic Press: London, UK, 1988; pp. 173–201.
46. Macura, S.; Farmer, B., II; Brown, L. An improved method for the determination of cross-relaxation rates from NOE data. *J. Magn. Reson.* **1986**, *70*, 493–499.
47. Hu, H.; Krishnamurthy, K. Revisiting the initial rate approximation in kinetic NOE measurements. *J. Magn. Reson.* **2006**, *182*, 173–177.
48. Haasnoot, C.A.G.; de Leeuw, F.A.A.M.; Altona, C. The relationship between proton-proton NMR coupling constants and substituents electronegativitis-I. *Tetrahedron* **1979**, *36*, 2783–2792.
49. Bally, T.; Rablen, P.R. Quantum-chemical simulation of 1H-NMR spectra. 2. Comparison of DFT-based procedures for computing proton-proton coupling constants in organic molecules. *J. Org. Chem.* **2011**, *76*, 4818–4830.
50. Bochevarov, A.; Harder, E.; Hughes, T.; Greenwood, J.; Braden, D.A.; Philipp, D.; Rinaldo, D.; Halls, M.; Zhang, J.; Friesner, R. Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences. *Int. J. Quantum Chem.* **2013**, *113*, 2110–2142.
51. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Scalmani, G.; Barone, V.; Menucci, B.; Petersson, G.A. *Gaussian 09,* Revision B.01; Available online: http://www.gaussian.com/ (accessed on 1 February 2014).

52. Reference methods for broth dilution antifungal susceptibility testing of yeasts; *Fourth International Supplement*; CLSI Document M27-S4; Clinical and Laboratory Standards Institute: Wayne, PA, USA, 2012.
53. Holm, D.K.; Petersen, L.M.; Klitgaard, A.; Knudsen, P.B.; Jarczynska, Z.D.; Nielsen, K.F.; Gotfredsen, C.H.; Larsen, T.O.; Mortensen, U.H. Molecular and chemical characterization of the biosynthesis of the 6-MSA derived meroterpenoid yanuthone D in *Aspergillus niger*. *Chem. Biol.* **2014**, *21*, 519–529.

*Sample Availability*: Not available.

Journal of Magnetic Resonance 263 (2016) 101–107

Contents lists available at ScienceDirect

Journal of Magnetic Resonance

journal homepage: www.elsevier.com/locate/jmr

# $S^3$ HMBC: Spin-State-Selective HMBC for accurate measurement of homonuclear coupling constants. Application to strychnine yielding thirteen hitherto unreported $J_{HH}$

Louise Kjaerulff [a,1], Andrew J. Benie [b], Casper Hoeck [a], Charlotte H. Gotfredsen [a,*], Ole W. Sørensen [a,*]

[a] Department of Chemistry, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
[b] Department of Peptide & Protein Biophysics, Novo Nordisk, 2760 Måløv, Denmark

ARTICLE INFO

*Article history:*
Received 1 August 2015
Revised 23 November 2015
Available online 6 January 2016

*Keywords:*
$S^3$ HMBC
HMBC
RDC
*J* coupling constants
Small molecules

ABSTRACT

A novel method, Spin-State-Selective ($S^3$) HMBC, for accurate measurement of homonuclear coupling constants is introduced. As characteristic for $S^3$ techniques, $S^3$ HMBC yields independent subspectra corresponding to particular passive spin states and thus allows determination of coupling constants between detected spins and homonuclear coupling partners along with relative signs. In the presented $S^3$ HMBC experiment, spin-state selection occurs via large one-bond coupling constants ensuring high editing accuracy and unequivocal sign determination of the homonuclear long-range relative to the associated one-bond coupling constant. The sensitivity of the new experiment is comparable to that of regular edited HMBC and the accuracy of the *J*/RDC measurement is as usual for E.COSY and $S^3$-type experiments independent of the size of the homonuclear coupling constant of interest. The merits of the method are demonstrated by an application to strychnine where thirteen $J_{HH}$ coupling constants not previously reported could be measured.



## 1. Introduction

*J* coupling constants have a long history of being among the most important parameters in structural assignment of small and large molecules by NMR spectroscopy. They provide angular information that can be used to determine stereochemical relationships. Particularly $^1$H–$^1$H coupling constants over three bonds are useful in this context due to the Karplus relation between $^3J_{HH}$ and the dihedral angle between the protons [1,2].

Over the past couple of decades *J* coupling constants (*J*s) have been supplemented by residual dipolar coupling constants (RDCs) in structure elucidation. RDCs are observable in weakly oriented solutions and are typically measured by the same methods as the *J*s. *J*s and RDCs can be both positive and negative and, unless the signs are known from other sources, it is crucial that they too are determined by the methods used to measure the size of the coupling constants.

Some of the techniques proposed for measurement of coupling constants rely on use of selective pulses which can be useful in special cases, but for general applications it is preferable to have techniques employing only non-selective pulses and thus not requiring prior detailed knowledge about spectral parameters for the molecules under study.

E.COSY-type techniques exploit simplified multidimensional multiplet patterns to extract even quite small coupling constants from cross peaks between two (active) spins when a third (passive) spin is coupled to both these spins, with a large coupling constant to at least one of the active spins. This is for example the principle behind the XLOC experiment measuring $^1$H–$^1$H coupling constants with one of the two protons being attached to a $^{13}$C spin, thus having a large one-bond coupling constant [3].

In E.COSY jargon the large coupling constant spreads out multiplet components in one of the dimensions allowing measurement of the small coupling constant of interest in an orthogonal dimension. Spin-State-Selective ($S^3$) techniques take an E.COSY multiplet simplification one step further by editing the spectrum into typically two independent subspectra corresponding to the two spin states of the passive spin.

The HMBC experiment [4] correlates e.g. $^1$H and $^{13}$C spins via the direct heteronuclear long-range coupling between the two spins and when another (passive) $^1$H spin is attached to the $^{13}$C

* Corresponding authors.
*E-mail addresses:* chg@kemi.dtu.dk (C.H. Gotfredsen), sorensen.ole.w@gmail.com (O.W. Sørensen).
[1] Present address: Department of Microbial Natural Products, Helmholtz Institute for Pharmaceutical Research Saarland, Helmholtz Centre for Infection Research and Pharmaceutical Biotechnology, Saarland University, 66123 Saarbrücken, Germany.

http://dx.doi.org/10.1016/j.jmr.2015.11.007
1090-7807/

spin that attached $^1$H spin can be subjected to S$^3$ editing via the large one-bond heteronuclear coupling thus allowing the spectrum of the active $^1$H to be split into two subspectra. That new experiment we dub S$^3$ HMBC and the necessary pulse sequences (outlined in Fig. 1) for its realization are derived below (*vide infra*).

## 2. The method

Spin-state selection can be done in a one-shot fashion where only particular spin states are excited, or in a way where both subspectra are generated from the same data set. The latter is usually advantageous when it comes to overall sensitivity and typically comprises addition and subtraction of data obtained with two different pulse sequences, possibly involving different phasing in the subspectra.

S$^3$ HMBC may be imagined as a simple version of two HMBC-type experiments differing only in one of them employing an additional S$^3$ $\pi_z$(S) pulse affecting only e.g. the resonances associated with α spin states of the attached proton(s). Then addition and subtraction of the two resulting data sets will yield two subspectra corresponding to the α and β spin states of the passive proton (I). The S$^3$ $\pi_z$(S) propagators can be written as:

$$e^{i\pi \mathrm{I}^{\alpha}\mathrm{S}_z} = e^{i\frac{\pi}{2}\mathrm{S}_z} e^{+i\frac{\pi}{2}2\mathrm{I}_z\mathrm{S}_z} \tag{1a}$$

$$e^{i\pi \mathrm{I}^{\beta}\mathrm{S}_z} = e^{i\frac{\pi}{2}\mathrm{S}_z} e^{-i\frac{\pi}{2}2\mathrm{I}_z\mathrm{S}_z} \tag{1b}$$

The initial $\pi/2$ S-spin $z$ rotation is of no great significance apart from generating a $\pi/2$ phase shift demanding different $t_1$ phase corrections for odd and even numbers of protons attached to the S spin (*vide infra*). However, these multiplicity-dependent $F_1$ phases do call for odd/even multiplicity editing to be incorporated into the S$^3$ HMBC experiment. One can say that each of the S-spin attached I spins adds a $\pi/2$ phase shift on top of the evolution under the coupling term.

The key feature of the propagators in Eq. (1) is the $\pi/2$ rotation under the heteronuclear (one-bond) coupling in a modified odd/even-edited HMBC experiment, which may then be linearly combined with a suitable standard odd/even-edited HMBC experiment. These additive and subtractive linear combinations produce the edited α and β subspectra, respectively, in both the odd and even multiplicity subspectra.

The chosen standard odd/even-edited HMBC experiment is the improved multiplicity-edited HMBC [5], as $J$ cross talk between the two subspectra in that experiment is a second-order effect in the deviation from the $J$ used in tuning the pertinent delays in the experiment.

Also a suitable modified odd/even-edited HMBC pulse sequence containing the additional $\pi/2$ rotation under the heteronuclear (one-bond) coupling is available in the literature, namely in the form of the HAT HMBC experiment [6]. That experiment was designed with the additional $\pi/2$ coupling rotation in order to attenuate cross peaks arising from heteronuclear couplings over more than two bonds and in that way distinguish HMBC correlations over two bonds from those over more bonds. HAT HMBC, where $J$ cross talk also is a second-order effect, is for general heteronuclear long-range correlation an alternative to the H2BC experiment [7,8].

With the two building blocks in place the derivation of the S$^3$ HMBC experiment amounts to combining data obtained with edited HMBC and HAT HMBC. For that to work a few amendments to the HAT HMBC pulse sequences must be made. Firstly it is necessary that the HAT HMBC sequences are extended, as to have the same delay between the $\pi/2$(H) pulse, the $\pi$(H) pulse and the start of acquisition as the edited HMBC sequences. This ensures equal evolution under $^1$H-spin chemical shifts and $^1$H–$^1$H homonuclear couplings. Furthermore, a decoupling period is added to ensure equal sample heating in all the sequences and the same evolution under heteronuclear couplings. Other corrective delays are negligible on the time scales of the pertinent coupling constants.

Both edited HMBC and S$^3$ HMBC are, for typical HMBC $\Delta$ delays, about 5–10% longer than the corresponding regular HMBC pulse sequence without these features. All of the individual pulse sequences of S$^3$ HMBC contain the same pulses as a regular HMBC.

S$^3$ HMBC spectra are constructed by first performing the odd/even or {CH/CH$_3$}/{C/CH$_2$} editing of the edited HMBC and HAT HMBC time domain data separately. Thereafter, the two even parts are added and subtracted from each other to produce the two S$^3$ HMBC spectra for measuring $J_{HH}$. Before performing the same procedure with the two odd parts of the standard HMBC and HAT HMBC data, the term $e^{i\frac{\pi}{2}S_z}$ in Eq. (1) must be considered because it effectively amounts to a $\pi/2$ phase shift for each attached proton in the $t_1$ dimension. Therefore, the odd HAT HMBC time domain data must be phase shifted by $\pi/2$ prior to linear combinations. This multiplication of the complex time domain data by $e^{-i\frac{\pi}{2}}$ amounts to swapping the real and imaginary data and changing the sign of the new imaginary data. The choice of $e^{-i\frac{\pi}{2}}$ rather than $e^{i\frac{\pi}{2}}$ is arbitrary, but it has the consequence of the center resonances from methylene groups and quaternary carbons appearing in the "all add" subspectrum of the linear combinations indicated in Eq. (2).

Disregarding the phase corrections, the S$^3$ HMBC linear combinations can with reference to the (*a*)–(*e*) pulse sequences in Fig. 1 be summarized as:

$$\left\{[a] \pm \frac{1}{2}[b+c]\right\} \pm \{d \pm e\} \tag{2}$$

The intra-brace operations are odd/even editing (HMBC and HAT HMBC) and the inter-brace operations lead to the S$^3$ editing.

Of these eight linear combinations in Eq. (2) the four with opposite signs in the first and last term are not relevant as they would combine odd and even data. The four final S$^3$ HMBC time-domain data sets are processed in the same way as regular HMBC data and all the spectra are presented in absolute-value mode.

The best sensitivity is obtained when the factor 1/2 in Eq. (2) is realized through adjustment of the number of scans for the (*b*) and (*c*) pulse sequences to be only half of those of the other pulse sequences. It can, of course, also be realized by numerical scaling of the data when sensitivity is not an issue.

One final small correction rooted in the high editing accuracy of the edited HMBC and HAT HMBC experiments can be taken into account. Effectively, the magnetization components in the (*b* and *c*) and (*d* and *e*) sequences that could give rise to out-of-phase error terms are eliminated in the linear combinations $[b+c]$ and $[d \pm e]$ of the data. Thus, except for τ perfectly matched to $(2(^1J_{CH}))^{-1}$, the intensities in these linear combinations will be slightly lower than those in the $[a]$ spectrum. Therefore, to reduce the already low level of $J$ cross talk even further, it is an option to scale $[b+c]$ and $[d \pm e]$ by small factors $k$ and $\sqrt{2k/(k+1)}$, respectively. These factors are applicable for CH groups and correspond to $-1/\cos(2\pi J\tau)$ and $1/\sin(\pi J\tau)$, respectively. This makes full elimination of $J$ cross talk on an individual cross peak basis possible, which, however, for the vast majority of applications is unnecessary. We typically employ a small correction of $k = 1.05$.

To make the S$^3$ HMBC experiment readily available we have written a Bruker pulse program acquiring the data in an interleaved pseudo-3D fashion and a Bruker AU program performing all the necessary linear combinations and phase corrections. The programs operate with four planes: plane #1 is the standard HMBC sequence in Fig. 1a, plane #2 is the linear combination of the sequences in Fig. 1b and c, and planes #3 and #4 are the HAT HMBC sequences in Fig. 1d and e, respectively. This AU program along with the pulse program is available from the website of

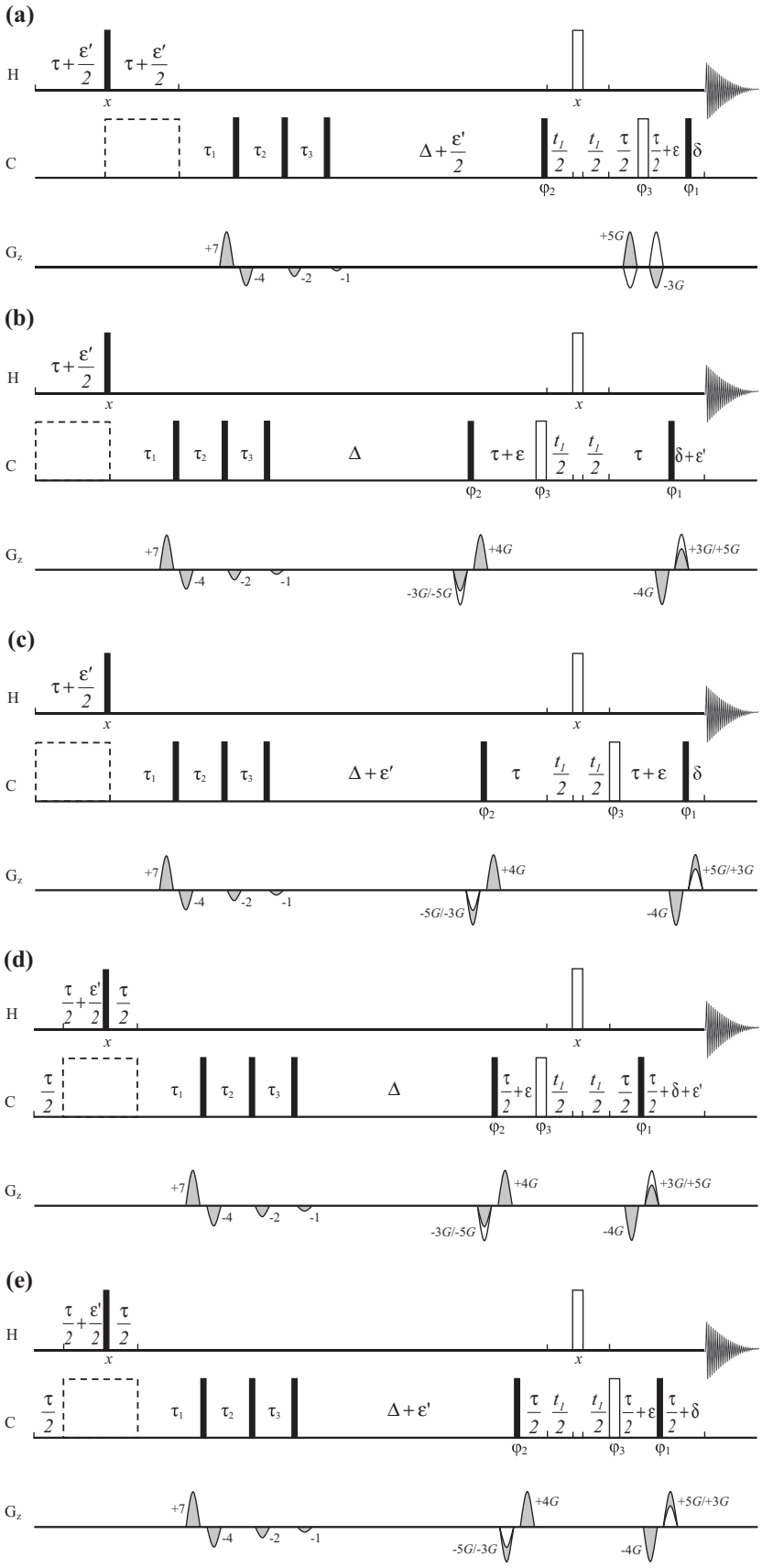



**Fig. 1.** $S^3$ HMBC pulse sequence comprising (a–c) the edited HMBC sequences and (d and e) the HAT HMBC sequences, all shown with a third-order low-pass *J* filter. Filled and open bars refer to $\pi/2$ and $\pi$ pulses, respectively, and the dashed open boxes represent $^{13}$C decoupling. $\tau = (2\ {}^1J_{CH})^{-1}$ or $({}^1J_{max} + {}^1J_{min})^{-1}$, $\delta$ is a gradient delay, $\varepsilon = 2\ t_{1/2,min} + t$ ($\pi_H$), $\varepsilon' = \varepsilon + t(\pi_C)$, $\tau_1 = ½[{}^1J_{min} + 0.07\ ({}^1J_{max} - {}^1J_{min})]^{-1}$, $\tau_2 = ({}^1J_{max} + {}^1J_{min})^{-1}$, $\tau_3 = ½[{}^1J_{max} - 0.07\ ({}^1J_{max} - {}^1J_{min})]^{-1}$. $\varphi_1 = \{x, -x, -x, x\}$, $\varphi_2 = \{x, x, 4(-x), x, x\}$, $\varphi_3 = \{4(x), 4(y), 4(-x), 4(-y)\}$. $\Delta$ is the delay for evolution under heteronuclear long-range couplings.

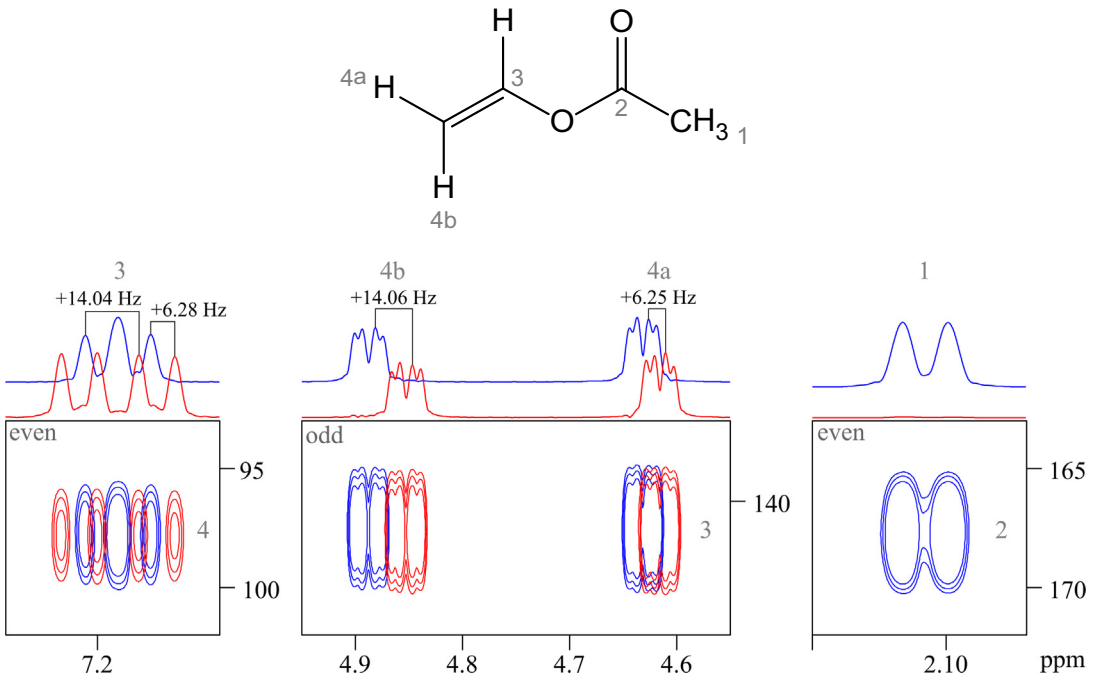



**Fig. 2.** Structure of vinyl acetate with atom numbering in gray and excerpts of cross peaks along with F1 sections from $S^3$ HMBC spectra of vinyl acetate (0.8 M in DMSO-$d_6$, 500 µl, 400 MHz, 16 scans, 4096 data points, 128 increments, zero filled to 65,536 × 1024, $J_{min}$ = 130 Hz, $J_{max}$ = 190 Hz, $\tau$ = 3.13 ms). The CH and $CH_2$ subspectra corresponding to multiplet components with an odd (even) number of $\alpha$ states for $^{13}$C-attached protons are plotted in blue (red). All subspectra and 1D sections were scaled to the same peak intensity. The two $^3J_{HH}$ coupling constants across the double bond were measured in two different multiplets.

the NMR group at DTU Chemistry and from Bruker's pulse program site.

## 3. Applications and results

For demonstration of $S^3$ HMBC two small molecules were chosen; vinyl acetate and strychnine. Vinyl acetate is a simple organic molecule with large $^3J_{HH}$ coupling constants and thus offers clear visual inspection of the level of cross talk [9,10]. Strychnine is one of the most widely used benchmark molecules for new small molecule NMR experiments, and most of the coupling constants have been measured and calculated [11–13], which makes it ideal for showing the potential of the $S^3$ HMBC experiment.

Four $S^3$ HMBC cross peaks of vinyl acetate are shown in Fig. 2 along with 1D sections exhibiting excellent editing accuracy, which translates into high accuracy measurement of the coupling constants.

In general, measurement of the coupling constant between the proton of a methine group with a $^{13}$C and a distant proton is straightforward and analogous to earlier E.COSY and $S^3$-type measurements. This is not the case for methylene and methyl groups, because the two $S^3$ HMBC subspectra are not just displacements of each other. Effectively, the two $S^3$ HMBC subspectra for methine and methylene carbons correspond to the multiplet components with odd or even numbers of $\alpha$ spin states for $^{13}$C-attached protons, respectively.

The *J*s in the $^{13}$C4(H4a,H4b)-H3 multiplet of vinyl acetate could not have been extracted from the unedited multiplet due to overlap of individual resonances, but in Fig. 2 it is possible, because the individual multiplet components are well resolved. This is, however, in general not fulfilled. As we shall see below, strychnine contains several methylene groups but most of them are not neighbor to another methylene group. Hence the $^3J_{HH}$ that cannot easily be determined from $^{13}CH_2$–$^{12}CH$ moieties can be determined from the complementary isotopomer, $^{12}CH_2$–$^{13}CH$.

Fig. 3 shows the structure of strychnine with atom numbering, while representative cross peak excerpts and 1D sections from $S^3$ HMBC subspectra are shown in Fig. 4.

34 $J_{HH}$ coupling constants in strychnine were determined by $S^3$ HMBC and compared to coupling constants from the literature [11–13] and DFT calculations [14]. The results are summarized in Table 1 while the details of the DFT calculations may be found in the Supplementary information.

Thirteen of the measured coupling constants had not previously been determined experimentally and could therefore only be compared to the results of the DFT calculation. The coupling constants measured using $S^3$ HMBC are in excellent agreement with both reference values and calculations (Table 1). H17ab is a second order multiplet system in the $^1$H NMR spectrum; however a coupling to H16 was measured. Since the chemical shifts of H17a and

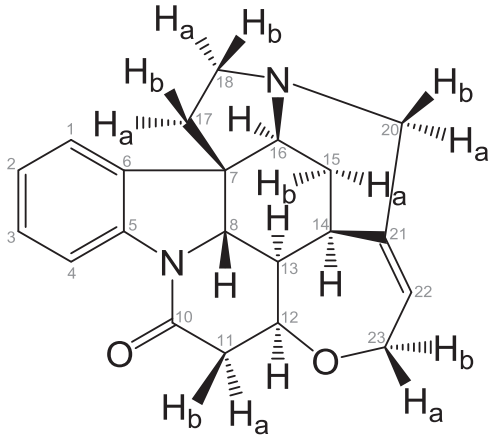



**Fig. 3.** Structure of strychnine with numbering and assignment of diastereotopic protons.

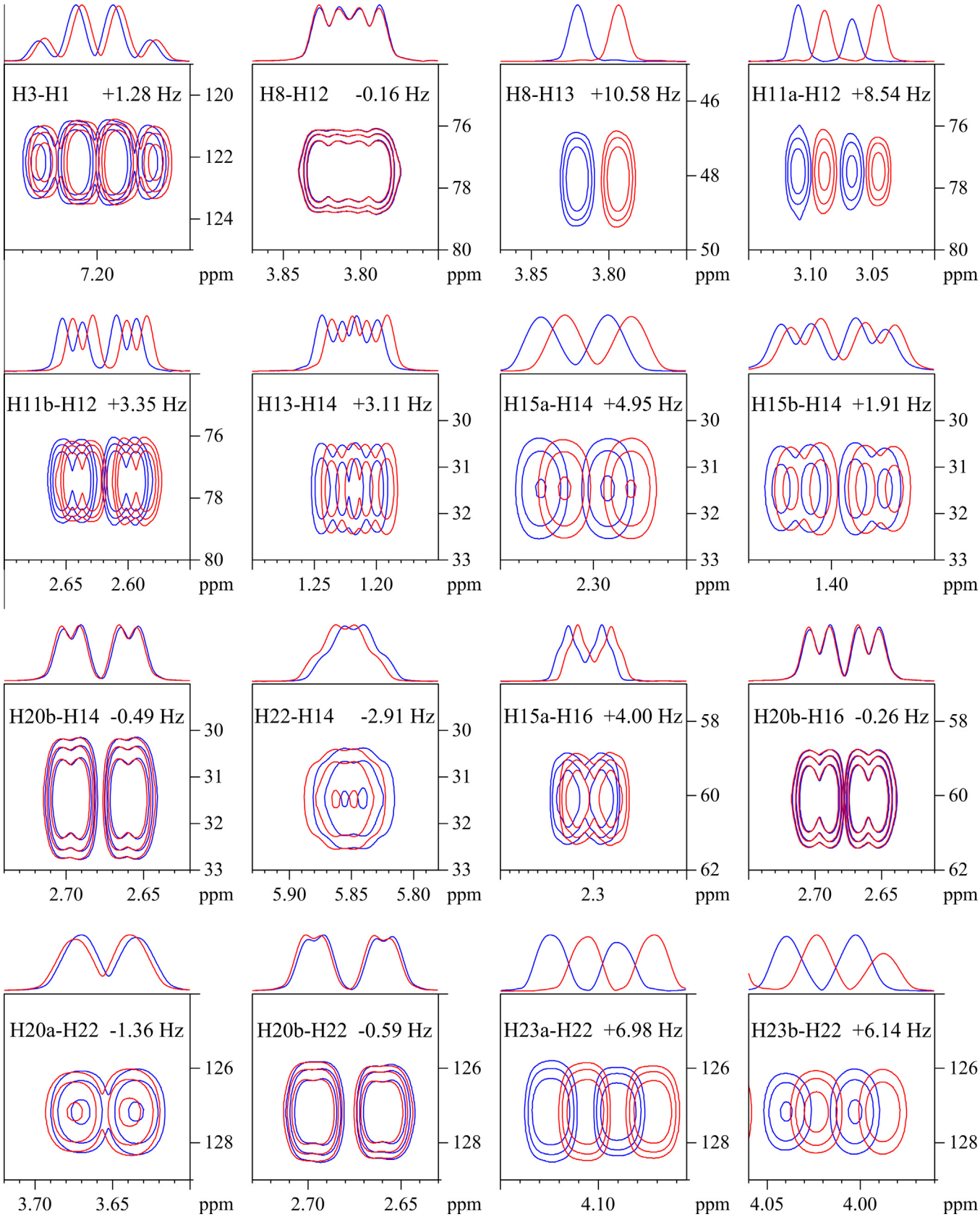



**Fig. 4.** $S^3$ HMBC cross peaks for strychnine (300 mM in $CDCl_3$, 500 µl, 400 MHz, 16 scans, 4096 data points, 256 increments, zero filled to 65,536 × 1024, standard HMBC delays ($J_{min}$ = 124 Hz, $J_{max}$ = 169 Hz, $\tau$ = 3.4 ms, delta = 65 ms). The coupling constants were measured by displacing 1D sections in analogy to the E.COSY method. The excerpts and 1D sections were scaled individually to the same intensity.

**Table 1**
Experimental and theoretical $J_{HH}$ coupling constants of strychnine measured at 400 MHz ([a] 800 MHz) [11–13]. Coupling constants extracted by a Matlab script with a minimum deviation of ±0.01 Hz. In some cases two values are reported, when $J$ can be measured in the cross peaks of both protons. Calculations were performed using B3LYP/6-31G(d,p)u + 1s (fermi contact only, linearly scaled by a factor of 0.9117). [b] H17ab was seen as a broad singlet and the measured coupling constant was taken as the sum. [c] Carter et al. [13] determined this coupling to 1.20 Hz, however this was suspected to be a mistake and therefore another reference was used for comparison [11]. [d] Top 30% of resonance used for coupling constant determination instead of the otherwise used 95% due to slight differences in the base of the peak. See detailed discussion in SI. The RMSD between the experimental and calculated coupling constants is 0.19–0.21 Hz based on the 400 MHz data, depending on the coupling constant used when two are measured (0.20 Hz if the average is used).

| #H1 | #H2 | Experimental $S^3$ HMBC (Hz) | Experimental $^1$H NMR (Hz) [13],*[11] | Calculated (Hz) |
|---|---|---|---|---|
| 1 | 2 | 7.68 | 7.41 | 7.61 |
| 1 | 3 | 1.50, 1.28 | 1.18 | 1.13 |
| 1 | 4 | 0.65, 0.66[d] | 0.45 | 0.55 |
| 2 | 4 | 1.24, 1.07 | 1.12 | 0.97 |
| 2 | 3 | 7.55[a] | 7.46 | 7.47 |
| 3 | 4 | 8.16 | 8.1 | 8.28 |
| 8 | 12 | −0.16, −0.01 | – | −0.10 |
| 8 | 13 | 10.58, 10.54[d] | 10.47 | 9.96 |
| 8 | 16 | −0.32, −0.20 | – | −0.27 |
| 11a | 12 | 8.54 | 8.44 | 8.43 |
| 11a | 13 | −0.46 | – | −0.31 |
| 11b | 12 | 3.35 | 3.28 | 3.75 |
| 12 | 13 | 3.23 | 3.12 | 3.58 |
| 12 | 14 | −0.32 | – | −0.20 |
| 12 | 15a | −0.13 | – | −0.09 |
| 12 | 23a | −0.10 | – | −0.09 |
| 12 | 23b | −0.16 | – | −0.02 |
| 13 | 14 | 3.11 | 3.10 | 3.27 |
| 13 | 15a | 0.27 | 0.49–0.59* | 0.38 |
| 13 | 15b | −0.40 | – | −0.28 |
| 14 | 15a | 4.95 | 4.58 | 4.60 |
| 14 | 15b | 1.91 | 1.98 | 1.93 |
| 14 | 16 | 0.65 | – | 0.54 |
| 14 | 20b | −0.49 | – | −0.50 |
| 14 | 22 | −2.91 | 2.82* | −3.03 |
| 15a | 16 | 4.00 | 4.11 | 3.89 |
| 15b | 16 | 2.15 | 1.82 | 2.02 |
| 16 | 17a | – | – | 0.04 |
| 16 | 17b | – | – | −0.18 |
| 16 | 17ab | −0.20[b] | – | −0.14 |
| 16 | 18a | −0.54 | – | −0.49 |
| 16 | 18b | 0.08 | – | 0.06 |
| 16 | 20b | −0.26 | – | −0.33 |
| 20a | 22 | −1.36 | 1.47* | −1.82 |
| 20b | 22 | −0.59 | *0.63*[c,*] | −0.58 |
| 22 | 23a | 6.98 | 6.88 | 6.86 |
| 22 | 23b | 6.14 | 5.74 | 6.14 |

H17b could not be discriminated only one correlation was observed. The measured coupling constant turned out to correspond well with the combined values of calculated coupling constants $J_{H16–H17a}$ and $J_{H16–H17b}$.

To eliminate any human error or over-fitting of the data, a Matlab script was used in order to extract the coupling constants directly from the spectra. The overall correlation between manually and automatically extracted couplings is good with an RMSD of 0.11 Hz. However some peaks are distorted due to strong coupling and automatic fitting is somewhat hampered. A comparison of couplings constants from manual versus automatic extraction and more details on the script can be found in the Supplementary information. Also a comparison of coupling constants from 400 and 800 MHz spectrometers is included.

Finally, it must be kept in mind that the current as well as practically all other similar techniques actually do not measure coupling constants but peak separations, but since they are identical in the weak coupling limit, and also for fair degrees of strong coupling, it is customary jargon to refer to them as techniques for measuring coupling constants. Thus slightly different values for the same coupling constant can be measured in different cross peaks, as there are a few examples of in Table 1. Based on measured peak separations (and intensities) it would take a fitting program to extract the actual coupling constants, something that is outside the scope of this paper.

## 4. Conclusions

A novel and robust NMR experiment, $S^3$ HMBC, has been introduced for accurate measurement of homonuclear coupling constants including their signs. For measurement of a $^{n+1}J_{HH}$ coupling constant, the critical parameter for the sensitivity is the heteronuclear long-range $^nJ_{CH}$ coupling constant on which all HMBC-type experiments are based. However, the accuracy of the measurement of $^{n+1}J_{HH}$ is independent of the size of that coupling constant.

The MBOB or broadband HMBC approach [15,16] for obtaining uniform excitation over a range of long-range $J_{CH}$ coupling constants is straightforward to implement in the $S^3$ HMBC experiment by recording data for different $\Delta$ values.

$S^3$ HMBC is expected to find widespread use for measurement of homonuclear $J$ and RDC coupling constants in small molecules, not least for analysis of carbohydrates where e.g. the small interglycosidic $^4J_{HH}$ coupling constants is expected to become visible via $S^3$ HMBC when the corresponding $^3J_{CH}$ coupling constants do not vanish [17].

## Acknowledgment

Funding from The Danish Council for Strategic Research is gratefully acknowledged.

## Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmr.2015.11.007.

## References


[1] M. Karplus, Contact electron-spin coupling of nuclear magnetic moments, J. Chem. Phys. 30 (1959) 11–15.
[2] M. Karplus, Vicinal proton coupling in nuclear magnetic resonance, J. Am. Chem. Soc. 85 (1963) 2870–2871.
[3] A. Meissner, O.W. Sørensen, Measurement of $J$(H, H) and long-range $J$(X, H) coupling constants in small molecules. Broadband XLOC and J-HMBC, Magn. Reson. Chem. 39 (2001) 49–52.
[4] A. Bax, M.F. Summers, $^{1}$H and $^{13}$C assignments from sensitivity-enhanced detection of heteronuclear multiple-bond connectivity by 2D multiple quantum NMR, J. Am. Chem. Soc. 108 (1986) 2093–2094.
[5] N. Nyberg, J.Ø. Duus, O.W. Sørensen, Heteronuclear two-bond correlation: suppressing heteronuclear three-bond or higher NMR correlations while enhancing two-bond correlations even for vanishing $^{2}J_{CH}$, J. Am. Chem. Soc. 127 (2005) 6154–6155.
[6] N. Nyberg, J.Ø. Duus, O.W. Sørensen, Editing of H2BC NMR spectra, Magn. Reson. Chem. 43 (2005) 971–974.
[7] A.J. Benie, O.W. Sørensen, Improved multiplicity-editing of HMBC NMR spectra, Magn. Reson. Chem. 44 (2006) 739–743.
[8] A.J. Benie, O.W. Sørensen, HAT HMBC: a hybrid of H2BC and HMBC overcoming shortcomings of both, J. Magn. Reson. 184 (2007) 315–321.
[9] K. Crecely, R. Crecely, J. Goldstein, $^{13}$C NMR spectra and long-range $^{13}$C–H couplings in substituted ethylenes, J. Mol. Spectrosc. 37 (1971) 252–259.
[10] P.E. Hansen, Carbon–hydrogen spin–spin coupling constants, Prog. NMR Spectrosc. 14 (1981) 175–296.
[11] J.C. Cobas, V. Constantino-Castillo, M. Martín-Pastor, F. del Río-Portilla, A two-stage approach to automatic determination of $^{1}$H NMR coupling constants, Magn. Reson. Chem. 43 (2005) 843–848.
[12] A. Bagno, F. Rastrelli, G. Saielli, Toward the complete prediction of the $^{1}$H and $^{13}$C NMR spectra of complex organic molecules by DFT methods: application to natural substances, Chemistry 12 (2006) 5514–5525.
[13] J.C. Carter, G.W. Luther, T.C. Long, Proton magnetic resonance spectra and assignments of strychnine and selectively deuterated strychnine, J. Magn. Reson. 15 (1974) 122–131.
[14] T. Bally, P.R. Rablen, Quantum-chemical simulation of $^{1}$H NMR spectra. 2. Comparison of DFT-based procedures for computing proton–proton coupling constants in organic molecules, J. Org. Chem. 76 (2011) 4818–4830.
[15] A. Meissner, O.W. Sørensen, Economizing spectrometer time and broadband excitation in small-molecule heteronuclear NMR correlation spectroscopy. Broadband HMBC, Magn. Reson. Chem. 38 (2000) 981–984.
[16] T. Schulte-Herbrüggen, A. Meissner, A. Papanikos, M. Meldal, O.W. Sørensen, Optimizing delays in the MBOB, broadband HMBC, and broadband XLOC NMR pulse sequences, J. Magn. Reson. 156 (2002) 282–294.
[17] A. Otter, D.R. Bundle, Long-range $^{4}J$ and $^{5}J$, including *inter*glycosidic correlations in gradient-enhanced homonuclear COSY experiments of oligosaccharides, J. Magn. Reson. Ser. B 109 (1995) 194–201.

ChemPubSoc Europe

DOI: 10.1002/chem.201600231

CHEMISTRY A European Journal Full Paper

Chemical Biology

# Specific Electrostatic Molecular Recognition in Water

Ming Li,[a] Casper Hoeck,[b] Sanne Schoffelen,[a] Charlotte H. Gotfredsen,[b] and Morten Meldal*[a]

**Abstract:** The identification of pairs of small peptides that recognize each other in water exclusively through electrostatic interactions is reported. The target peptide and a structure-biased combinatorial ligand library consisting of ≈78 125 compounds were synthesized on different sized beads. Peptide–peptide interactions could conveniently be observed by clustering of the small, fluorescently labeled target beads on the surface of larger ligand-containing beads. Sequences of isolated hits were determined by MS/MS. The interactions of the complex showing the highest affinity were investigated by a novel single-bead binding assay and by 2D NMR spectroscopy. Molecular dynamics (MD) studies revealed a putative mode of interaction for this unusual electrostatic binding event. High binding specificity occurred through a combination of topological matching and electrostatic and hydrogen-bond complementarities. From MD simulations binding also seemed to involve three tightly bound water molecules in the interface between the binding partners. Binding constants in the submicromolar range, useful for biomolecular adhesion and in nanostructure design, were measured.

## Introduction

Molecular recognition plays a central role in biology and is essential for life.[1] It is a crucial aspect of many biological processes, such as protein folding,[2] the structural organization of cells and organelles, signal transduction,[3] and the immune response.[4] The phenomenon is defined as the specific interaction between two or more molecules, mediated by non-covalent forces, including hydrogen bonding, hydrophobic forces, electrostatic and van der Waals interactions, aromatic stacking, and metal coordination.[5] It is a complex and multifaceted process, which depends not only on the structure and functional groups of the molecules involved, but also on the environment. In particular, in water, parameters such as pH, salt concentration, and temperature, have a large effect.[2]

In fact, the most prevailing means of cellular signaling and event control in biology is constituted by the recognition between a peptide and a protein. Therefore, understanding this process is a key element in the development of drugs[6] whose mode of action is to interfere with such peptide–protein interactions. Biomolecular binding events are characterized by binding constants with values ranging from weak[7] to intermediate[8,9] and strong.[10] For tight binding of peptides to receptors, typically occurring in an aqueous environment, hydrophobic interactions are the principal thermodynamic driving force. During the binding event, entropy is gained owing to the release of organized water from the binding interface. This, in turn, enables the establishment of a hydrogen-bonding network, electrostatic interactions, and salt bridges, which together contribute to the enthalpy of binding.[11–13] In protein–peptide interactions, the final hydrogen-bond network frequently includes residual tightly bound water molecules.

Specific biomolecular interactions have been mimicked, both by design[14] and combinatorial screening.[15] In material sciences, supramolecular materials based on non-covalent self-association have been developed involving, for example, cucurbit[7]uril and biotin[6]uril.[16–18] In these examples, recognition is of the lock-and-key type, which differs significantly from the induced fit of designed synthetic peptide–receptor pairs.[19–21] Artificial receptors developed through combinatorial screening show binding constants in the range $10^{-5}$–$10^{-3}$ M, and include examples such as tweezer- and tripodal peptides,[22] bicyclic peptide receptors,[23] and peptidomimetic receptors binding to a Ras-protein. Interestingly, Maeda et al. reported the identification of pentapeptides binding to a dopamine D2 receptor fragment with $K_d$ values of $10^{-5}$ M by using a combinatorial bead–bead interaction assay.[24]

Elucidating the mechanism of molecular recognition is challenging and may be best addressed through selection processes and molecular evolution. In an effort to increase our understanding of molecular recognition, we aim to systematically investigate each of the major types of interaction separately, starting with electrostatic interactions. For this purpose, we have developed a bead-based screening protocol that, in

[a] M. Li, Dr. S. Schoffelen, Prof. M. Meldal
Department of Chemistry, Center for Evolutionary Chemical Biology
University of Copenhagen, Universitetsparken 5
2100, Copenhagen (Denmark)
E-mail: Meldal@nano.ku.dk

[b] C. Hoeck, Dr. C. H. Gotfredsen
Department of Chemistry, Technical University of Denmark
Building 207, 2800, Kongens Lyngby (Denmark)

Supporting information and ORCID from the author for this article are available on the WWW under http://dx.doi.org/10.1002/chem.201600231.

principle, allows for up to $10^{12}$ peptide–peptide interactions to be evaluated simultaneously. In the present paper, we describe this methodology as well as a model study performed to identify pairs of small linear polar peptides that specifically recognize each other in an aqueous environment. The specific, complex, and multifaceted interaction between one of the identified interacting peptide pairs, displaying a binding constant in the submicromolar range, was investigated by 2D NMR spectroscopy and molecular dynamics.

## Results and Discussion

We have previously identified specific recognitions between adhesion molecules and the lipid bilayer of cells through combinatorial screening.[25] Based on this study, we suggest that it should be feasible to find pairs of peptides that specifically bind to each other by using a similar approach. In contrast to identification of receptor–ligand interactions by design, the identification through combinatorial screening can accommodate, within the library design, all possible combinations of hydrogen bonding, electrostatic, and hydrophobic interactions, as well as the conformational space of the potentially interacting molecules.

As a model for this screen, the polar decapeptide QTRTNTHRDG was selected and screened for binding to a combinatorial library of 78125 octapeptides. The target compound was immobilized on 100 μm microbeads labeled with a fluorescent dye, whereas the octapeptide ligand library was synthesized on non-labeled 400 μm macrobeads. This provided a convenient way of identifying positive hits by bead-to-bead adherence, as depicted in Figure 1.

### Design and synthesis of target peptide and ligand library

The target sequence and the ligand library were synthesized on $PEGA_{800}$ and $PEGA_{1900}$ resins, respectively. Poly[acryloyl-bis (aminopropyl)polyethylene glycol] (PEGA) resin was selected as the solid support because of its low background fluorescence

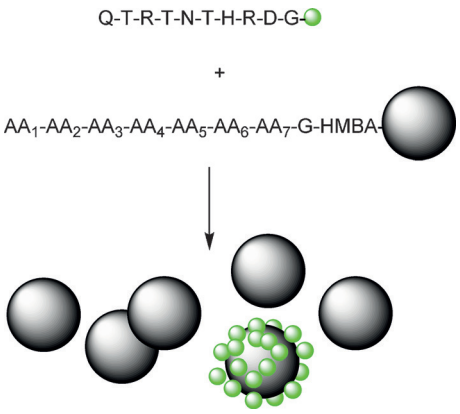



**Figure 1.** Schematic representation of the bead-based screening protocol employed in this study to identify peptide–peptide interactions. The target peptide was synthesized on fluorescently labeled, 100 μm-sized microbeads. The ligand library was synthesized by split/mix assembly on non-fluorescent 400 μm-sized macrobeads.

and absence of unspecific interactions with reagents, peptides, and proteins. Moreover, PEGA shows excellent swelling properties in polar solvents and is generally biocompatible.[23–25]

Prior to library assembly, the target was modelled in Molecular Operating Environment (MOE™) and virtually screened for possible recognition elements. We analyzed substitution patterns showing potential interactions with the selected target peptide, using an octapeptide structure to generate ligands with increased probability of strong and specific interactions. The C-terminal position was maintained as glycine, as the absence of a side-chain in glycine allows for fast and quantitative coupling and cleavage from the resin. For the other seven positions, five different amino acids, both proteinogenic and non-proteinogenic, were used, as presented in Table 1.

**Table 1.** The composition of the split/mix ligand library.

| 1[a] | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|
| D | D | $G_F$[b] | E | Dab | D | D | G |
| E | E | K | H | Dap | E | E | |
| L | H | O | R | O | F | F | |
| F | N | R | O | R | L | L | |
| N | Q | Y | Q | Q | N | N | |

[a] Numbers indicate the position in the octapeptide starting from the N-terminus. [b] O, L-ornithine; $G_F$, guanidino-L-phenylalanine; Dab, L-2,4-diaminobutyric acid; Dap, L-2,3-diaminopropioic acid.

Solid-phase synthesis was performed by using a fluorenylmethyloxycarbonyl (Fmoc)-based protocol as shown in Scheme S1 (in the Supporting Information). The ligand library was synthesized by using the base-labile linker 4-(hydroxymethyl)benzamide (HMBA). In this way, peptides could be deprotected on the solid support and cleaved upon hit identification for MS/MS analysis. The one-bead one-compound (OBOC) library (78125 members) was prepared by the solid-phase split-and-combine strategy on >100000 beads, thus providing the majority of the theoretically possible sequences.

The small $PEGA_{800}$ particles were labeled with ATOTA (tris(dialkylamino)-trioxatriangulenium dye) by coupling a mixture of ATOTA-OSu and Fmoc-Gly-OH ($n/n = 1:9$). The ATOTA fluorophore was selected based on its chemical stability under both basic and acidic conditions.[26] The target peptide was synthesized on the fluorescently labeled microbeads thus obtained. No linker was added between the solid support and peptide, such that the target molecule would stay attached to the microbeads during cleavage of ligands from the macrobead.

### Selection of positive hits

Combinatorial screening of ligands for proteins has conventionally been performed by labeling the protein target with biotin or fluorescent labels prior to incubation with the ligand library.[15] Based on our success of adhering cells to beads, we anticipated that bead–bead interaction as a measure of binding could have a range of advantages in the screening process.

The present screen directly relates to the property we were looking for, thus reducing false positives. It also allows us to distinguish strong and weak binders by shear force, and it facilitates binding competition studies with hydrogen-bond donors and acceptors or electrostatic competitors. Key to its success is the easy identification and isolation of macrobeads that are covered with microbeads. Non-interacting microbeads can be removed by sieving, after which, fluorescence sorting can be employed to select positive hits.

When the macrobead library was mixed with the target-containing microbeads, bead-to-bead adherence was observed (Figure 2). The number of microbeads clustered to the individual macrobeads was found to vary, suggesting a variation in the degree of adherence and thus varying affinities between the ligands and target peptide. Several macrobeads were completely covered with microbeads, indicating maximum adhesion. By combining bead-to-bead adherence with the fluorescent signal provided by the microbeads, it was possible to track and isolate positive hits by using fluorescence microscopy. At this stage, 33 initial hits were selected from the 78125-member ligand library.

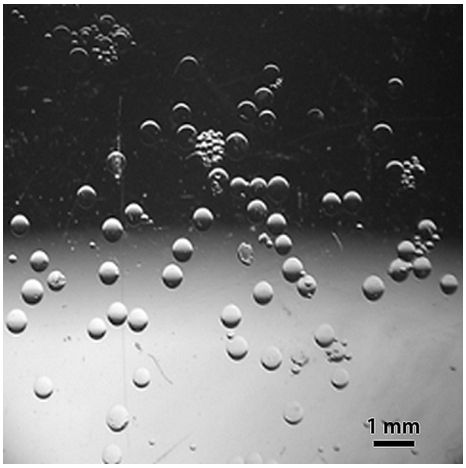



**Figure 2.** Microscopy image showing assembly of target-containing microbeads clustered around a macrobead.

As a pre-assessment for the degree of affinity, the isolated bead clusters were incubated in the presence of salt by using an aqueous solution of 50 mM phosphate buffered saline (PBS, pH 7.4).Some of the hits were found to release the microbeads under these conditions, implying weak interaction. In the end, only 16 hits were obtained from 78125 possible entities ($P = 0.02\% < 0.5\%$), indicating very specific recognition between the target peptide and library-derived ligands.

## Peptide sequence identification by MS/MS

The 16 isolated hits were subjected to sequence analysis. Traditionally, Edman degradation has been used in OBOC combinatorial chemistry. However, rapid and straightforward HR-MS/MS has recently been shown to facilitate high-throughput analysis of the limited (nmol) quantities of compound available from single library beads.[27,28] We found that matrix-assisted laser desorption/ionization Fourier-transform ion cyclotron resonance (MALDI-FT-ICR) mass spectrometry was extremely sensitive and provided accurate sequence information of pmol quantities of peptide by implementation of a variety of fragmentation conditions.

Ligands were released from isolated beads by using a small volume of triethylamine (TEA) in water. For hydrophilic sequences in particular, TEA/water was superior to the conventional NaOH cleavage in providing sodium-free spectra for MS/MS. The sequence data were analyzed with the software BioTools™ (Bruker Daltonics) and the custom-made LibMSCalc. The latter takes advantage of the knowledge of the library components, thereby greatly improving fidelity, and is available from the authors.

As a representative example, the MS/MS spectrum of peptide **7** is shown in Figure 3. The sequences obtained from the library-derived ligands **1**–**16** are presented in Table 2. In addition, Table 2 contains compounds **17**, **18**, and **19**. These

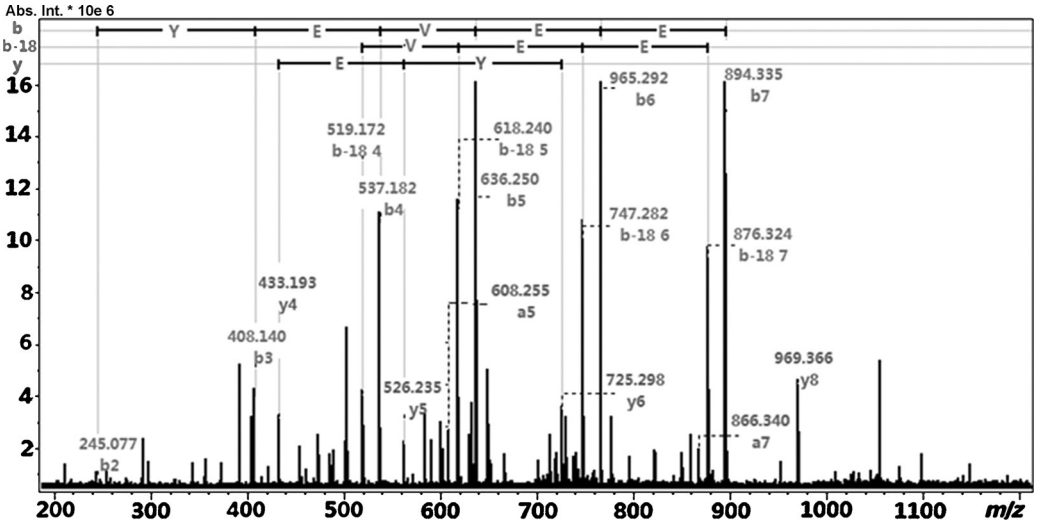



**Figure 3.** MS/MS sequence determined for compound **7**. Peaks belonging to b, y and a ion series are assigned; b-18 is a b ion that has lost a water molecule.

**Table 2.** Library-derived ligands and their affinity for targeting QTRTNTHRDG under various conditions.

| Cmpd | Sequence | $[M+H]^+$ | Milli-Q[a] | Urea | FBS |
|---|---|---|---|---|---|
| **1** | ENYQRFFG | 1060.4864 | + | – | + |
| **2** | EDG$_F$QDapFDG | 952.3985 | – | ++ | ++ |
| **3** | DEG$_F$EVDDG | 982.3744 | +++ | + | + |
| **4** | EDYEREDG | 1012.3840 | +++ | + | + |
| **5** | EQYEDabDDG | 955.3636 | ++ | ++ | + |
| **6** | DDG$_F$HDapDDG | 963.3190 | +++ | ++ | + |
| **7** | EDYEVEEG | 969.3687 | ++++ | +++ | +++ |
| **8** | DQYERDDG | 997.3845 | ++++ | + | + |
| **9** | DEYRVEEG | 996.4266 | +++ | – | + |
| **10** | EDG$_F$EOEEG | 1025.4171 | ++++ | + | + |
| **11** | EDYQVEEG | 968.3835 | +++ | ++ | + |
| **12** | EHOHOFLG | 967.5116 | + | ++ | + |
| **13** | EDG$_F$EDabEDG | 997.3846 | +++ | ++ | ++ |
| **14** | DDYEODDG | 942.3310 | ++++ | ++ | ++ |
| **15** | EDREVDEG | 948.3904 | ++++ | +++ | + |
| **16** | EDG$_F$EVDDG | 982.3755 | ++++ | +++ | + |
| **17** | EDDWDDG | 851.7515 | ++++ | +++ | +++ |
| **18** | EDYEVDDG | 941.8740 | +++ | +++ | ++ |
| **19** | EDYEWEEG | 1057.0169 | ++++ | +++ | +++ |

[a] The strength of binding of target-containing microbeads to ligands on macrobeads in water, 0.4 M urea, and 10% fetal bovine serum (FBS) as indicated in a qualitative manner. O, L-ornithine; G$_F$, guanidino-L-phenylalanine; Dab, L-2,4-diaminobutyric acid; Dap, L-2,3-diaminopropioic acid.

peptides were designed based on the level of activity of ligands **1**–**16** and will be discussed in more detail below.

### Binding studies under competitive conditions

The target sequence was deliberately designed to avoid unspecific hydrophobic interactions. Therefore, the ligand–target interaction was expected to be dominated by hydrogen-bonding and charge–charge interactions. To confirm whether this was the case, the hits were resynthesized and subjected to more rigorous screening. The results of this qualitative study are presented in Table 2.

Most of the molecules still presented high affinity in milli-Q water. However, in an aqueous solution of 0.4 M NaCl, binding was completely abolished, probably as a result of reduced charge–charge interactions in the presence of $Na^+$ and $Cl^-$ ions. In contrast, when the beads were mixed in water prior to addition of NaCl, adhesion between beads was surprisingly maintained. This indicated that on the macroscopic level charge–charge interactions were a required long-range interaction ($>1$ μm) for the initial bead attraction, but that the actual strong recognition was mediated by a variety of short-range hydrogen-bonding, electrostatic, and polar van der Waals interactions.

To investigate the significance of hydrogen bonds, the interacting beads were incubated in 0.4 M urea. Except for ligands **1** and **9**, the microbeads were still capable of assembling onto the macrobeads. However, the interaction was weakened, confirming that hydrogen bonding played a role. To mimic more biologically relevant conditions, the beads were incubated in 10% fetal bovine serum (FBS), frequently used in the study of biomolecular interactions. All of the ligand peptides presented recognition of the target under these conditions. Peptide **7** (EDYEVEEG) was found to be superior to all other library-derived ligands.

The results obtained from the screening indicated two regions of electrostatic interaction in positions $Aa_1Aa_2$ and $Aa_6Aa_7$ of the octapeptide. The present assay in which hits are identified by bead–bead adherence may well be biased towards weak charge–charge interactions, known to be active over longer distances. Comparison of the intramolecular NOE's of the individual compounds and those of the complex indicated that the well-defined target structure, spatial preorganization of the charges, and additional features for hydrogen bonding secured strong binding. Although, the 16 identified ligands were very similar, the degree of adherence varied significantly. The unique motif and few hits identified both indicate that the overall recognition is quite specific.

### Determination of binding constants

It was decided to determine the binding constants of the four most promising compounds from Table 2, that is, ligand **7**, **17**, **18**, and **19**. The binding constants were determined by using a novel binding assay based on single beads fixed in a microchannel of a custom-made flow cell (see Figure S1 in the Supporting Information). Binding is measured by fluorescence microscopy, visualizing the capture of fluorescently labeled binding partners to ligand molecules, immobilized by amino-functionalized PEG tentacles in the bead. The binding assay takes advantage of the stealth properties of the $PEGA_{1900}$ resin. This resin consists of a network of PEG chains and interaction with proteins and peptides is negligible. The resin swells 15–20-fold in buffer solutions, thereby providing extremely fast diffusion of proteins up to 60 kDa.[29]

The binding assay is particularly useful for the measurement of small-molecule interactions where surface plasmon resonance (SPR) can be a challenge. Binding constants for bimolecular interactions can in principle be determined by only two background-corrected measurements of fluorescence intensities measured at saturation ($I_{F\,max}$) and at intermediate binding concentration ($I_F$ at $[T]_0$) of the target in solution by using the equation $K_d = [T]_0 \times I_{F\,max}/I_F - [T]_0$ (see the Supporting Information). Accurate determinations are obtained by fitting to a complete binding curve.

The target was labeled with fluorescein and a dilution series from 1 nM to 100 μM was prepared. The single bead was labeled with ligand and incubated with different concentrations of target in solution. First, the fluorescence intensity at point of saturation was determined with the second highest and highest concentrations of target. Subsequently, the labeled target was eluted from the bead to verify reversible binding. Finally, the bead was incubated with the labeled target solution from low to high concentration, recording the fluorescence intensity of the bead in the same region of interest.

The binding curve for ligand **7** is presented in Figure 4 together with the background curve representing non-specific binding (acetamide-modified PEGA bead). The binding data

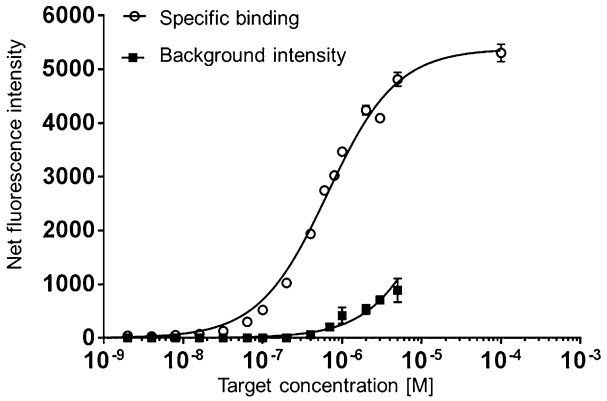



**Figure 4.** Net fluorescence intensity versus concentration of ligand **7** in solution. Squares: background, open circles: intensity obtained with compound **7** upon subtraction of the background.

**Table 3.** Binding constants, $K_d$, for binding of target **20** to the four superior ligands identified.

| Ligand | Structure | $K_d$ [μM] |
|---|---|---|
| **7** | EDYEVEEG | 0.658 ± 0.029 |
| **17** | EDDWDDG | 3.37 ± 0.25 |
| **18** | EDYEVDDG | 6.95 ± 1.49 |
| **19** | EDYEWEEG | 0.661 ± 0.059 |

were fitted [Eq. (8) in the Supporting Information] providing the $K_d$ values presented in Table 3. The binding curves for ligand **17**, **18**, and **19** are presented in Figure SI2 (in the Supporting Information).

Ligand **7** (EDYEVEEG) had the highest affinity with a $K_d$ of 0.658(±0.029) μM. Ligand **18** (EDYEVDDG), which was generated from the consensus of the library screening, only showed weak affinity during the subsequent binding study. This indicates that the unique interactions required for specificity are best identified by direct selection from a library. The extent of the preorganized, three-dimensional structure could play a role in the recognition and the structures were further studied by NMR spectroscopy.

### Characterization of the ligand–target interaction by NMR spectroscopy

Further investigation of the interaction of ligand **7** (EDYEVEEG) and target **20** (QTRTNTHRDG, Figure 5) was performed by 1D and 2D NMR spectroscopy as detailed in the Supporting Information.

Nuclear Overhauser effect (NOE) spectroscopy was the initial focus owing to the through-space correlations and the possibility to obtain distances between nuclei. Although several intra-molecular NOEs were identified and subsequently quantified, no inter-molecular NOEs were observed to support the interaction of the two peptides in the mixture. This may in part be attributed to a crowdedness of the sidechain region of the spectra, and most likely the absence of inter-peptide backbone contacts.

However, other clear NMR indications of complex formation were observed. There was a large shift in the rotational correlation time. This change was visible in the NOESY spectra, as the ratio of cross-peak to diagonal-peak intensity was significantly higher for the mix of the peptides compared with the single peptides. An increase in the size of the peptide, which should correlate to a longer correlation time, would lead to an increase in the NOE intensities (Tables SI1 and SI2 in the Supporting Information). Moreover, a number of variations in specific inter-proton distances from NOEs of the target (largest for residue R3) resulting from induced conformational changes

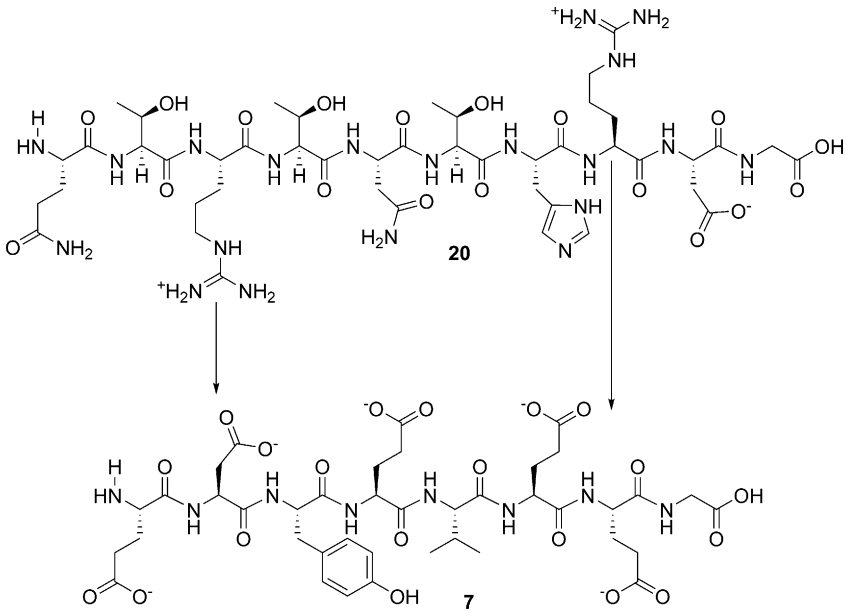



**Figure 5.** The superior recognition pair consisting of target **20** and ligand **7**. Screening was performed at pH 7.0, hence, R carried a positive charge while E and D were negative. Arrows indicate the important charge–charge interactions.

suggested significant interaction of ligand and target in the mixture.

To further evaluate the binding interactions, we investigated the change of backbone chemical shift (Figure 6). The overall change in chemical shifts was insignificant in the target and small in the ligand, indicating that binding occurred essentially without major conformational changes. From the change in backbone chemical shifts for ligand residues **3**, **4**, **5**, **6**, and **7**, it can be concluded that the ligand undergoes small adaptations to the target structure during the binding event, possibly as the two ends interact with the arginine residues of the target, which induce a conformational change in the central part of the ligand and an additional interaction with the target. Titration studies (see Figures SI3–SI5 in the Supporting Information) confirmed that the changes in chemical shift are gradual when titrating target **20** into the ligand **7**. The amino-terminal alfa protons in both ligand **7** and target **20** were particularly sensitive to peptide interaction (Figure SI5). Dilution of the mixture returned the chemical shift values back to the chemical shift values of the free ligand. However, the chemical shifts for the free ligand were not reached at concentrations suitable for NMR ($>\approx 10^{-5}$ M, compare Figure SI6 with SI4 in the Supporting Information).

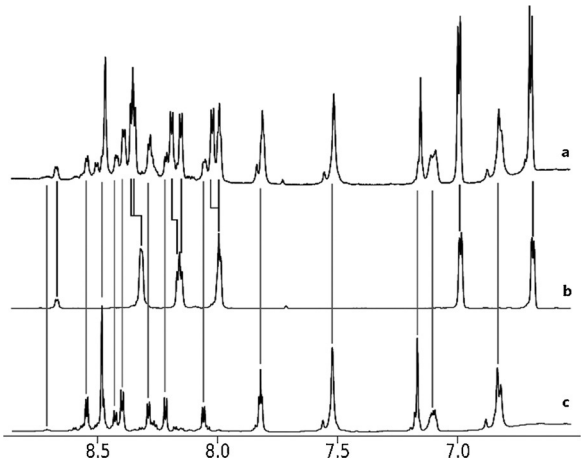



**Figure 6.** Changes in the chemical shifts in the amide region upon complexation (a) of ligand **7** (b) and target **20** (c).

### Restrained molecular dynamics (MD)

No direct NOEs could be observed between target **20** and ligand **7**. To provide structure information on distances, NOEs and MD calculations of the complex were combined. The NOE distances derived from the complex of ligand **7** and target **20**, were larger than those of the free compounds owing to an increase in cross relaxation rate, leading to a better signal-to-noise ratio. Therefore, these NOE determinations were first implemented in annealing MD calculations for each individual compound.

Structures that fulfilled all the observed distances were readily obtained (see the Supporting Information). Assuming that charge–charge interactions of both R3 and R8 are involved and minimal structural changes occur during binding, only four orientations of the ligand and target were feasible. Two of those were most probable due to the NOEs of Arg-NH to Arg side chain protons. The NOE restraints were maintained while MD calculations of the complex in these two orientations were performed, including several rounds of annealing from 700 K. The complex presented in Figure 7 was by far the best fit and in addition provided the highest calculated binding energy. The free ligand conformation was determined from those of the complex by implementing the NOE restraints determined for the individual compounds in MD calculations with annealing (see Figure SI8 c and e in the Supporting Information).

The complex provided turns both in the ligand and the target sequence. The ligand had a turn at residue E5 and the target at residues T6-H7, in strong agreement with the ob-

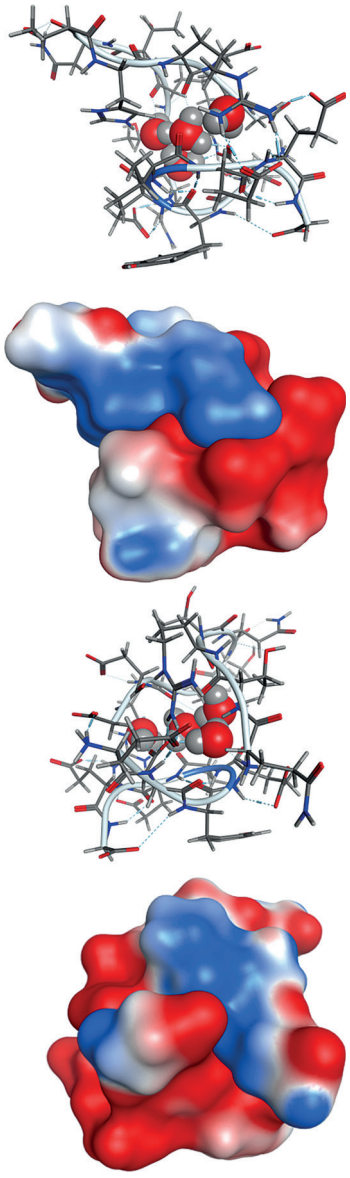


**Figure 7.** Model of the **7**–**20** complex derived through restrained (88 distances) MD computer simulations, showing the ball-and-stick representation and Connolly molecular surface from two different angles.

served NOEs. The His side chain folded into the turn and showed a significant number of NOEs to neighboring residues. The backbone NOE distances suggested a high degree of β-strand structure from short $NH^n$ to $C_\alpha H^{n-1}$ distances, which were in agreement with the optimized structures, as illustrated in separate images of the structures in the Supporting Information. In the complex, there is tight interaction between target residue R3 and ligand residues E1 and D2, whereas target residue R8 is suspended between the carboxylates of ligand residues E6 and E7.

The complex presents rim contact with a closed cavity containing four water molecules (Figure 7), three of which are in tight hydrogen-bonding contact with both ligand residue D2 and target residues Q1, T2, R3, H7, and R8 (see Figure SI9 in the Supporting Information). We assume that this could be one reason for the fact that no NOEs are observed between the ligand and target in spite of the strong binding. The intermolecular distances in the structure derived for the complex did not give rise to expected intermolecular NOEs. Most importantly, the ligand and target presented surface complementarity in the topology of the complex, and the conformations of the free compounds are quite similar to those observed in the complex in full agreement with the NMR data. Although the structure was an optimized fit to the observed NMR-data, we acknowledge that this is a putative structure in the absence of crystal data and direct intermolecular NOEs for this novel type of binding.

## Conclusions

In the present work, highly specific recognition based exclusively on electrostatic interactions between linear peptides in water was achieved through combinatorial screening. As demonstrated, interaction between two peptides does not have to involve lipophilic domains, but can be based entirely on the optimal spatial arrangement of functional groups involved in charge–charge and hydrogen-bond interactions. This seems to be dependent on the native topology of the binding partners and their topological matching of complementary electrostatics. An important observation was the fact that specific molecular interaction does not always provide NOEs resulting from short proton–proton distances, although tumbling and other spectral parameters clearly show the formation of a binary complex. Particularly interesting was the contribution of bound water for the complex formation suggested by restrained MD calculations. Such pairs of target peptides and matching ligands may find many applications for selective recognition events in protein purification and in cell biology. We presented a peptide–target-based assay for finding the optimal recognition pair by using a library biased towards the target sequence. Other targets would have called for different bias and would probably result in different modes of recognition. Such studies are in progress.

## Experimental Section

### Materials

Beaded poly-ethylene glycol-poly-dimethyl acrylamide copolymer $PEGA_{800}$ resin was synthesized by radical inverse suspension polymerization of bis-acrylamido isopropyl polyethylene glycol (900) and dimethyl acrylamide (5% by weight), as previously described.[29] $PEGA_{1900}$ was prepared accordingly by using bis-acrylamido isopropyl polyethylene glycol (2000) and acrylamide (5% by weight). Fmoc amino acids were obtained from Bachem AG or from Chemimpex. Solvents were high-purity grade from Sigma–Aldrich and were used without further purification. Milli-Q water (Millipore) was used throughout the work.

### Synthesis of target on resin without HMBA linker

Dry $PEGA_{800}$ resin (1.0 g, 0.33–0.35 mmol) was swelled in $CH_2Cl_2$ and washed with DMF (3 vol). The resin was coupled for 2 h with a mixture of Fmoc-Gly-OH (272.9 mg, 0.918 mmol) and ATOTA-OSu (59.5 mg, 0.102 mmol) pre-activated with 1*H*-benzotriazolyl tetramethyluronium tetrafluoroborate (TBTU, 314.4 mg, 0.98 mmol) and 4-ethylmorpholine (NEM, 173 μL, 1.36 mmol) in DMF (10 mL). Prior to the coupling, the fluorescence of the resin was confirmed by fluorescence microscopy and the resin presented strong fluorescence at 520 nm. Removal of the Fmoc-group with 20% piperidine in DMF was confirmed by the Kaiser test, which showed intensely colored blue beads. The subsequent Fmoc amino acids were coupled to the resin by following the general solid-phase peptide synthesis (SPPS) procedure (see the Supporting Information).

### Library synthesis

A peptide library was synthesized on dried $PEGA_{1900}$ resin (1.0 g, bead diameter 400–500 μm, 0.2 mmol $g^{-1}$, $\approx 1.1\times10^5$ beads). Before transfer into a 20-well multiple column peptide synthesizer (MCPS), the HMBA-modified $PEGA_{1900}$ resin was coupled with Fmoc-Gly-OH in a sinter-fitted syringe.[30] The resin was transferred to the MCPS instrument. To equally divide the beads into the columns, the MPCS was filled with DMF (45 mL), sealed, turned over, and shaken for 10 min. Then, the MCPS was turned upright and an even distribution of the resin was visually confirmed. The Fmoc group was removed by 20% piperidine in DMF including 2% HOBt. After washing with DMF (6 vol), Fmoc-Asn(trt)-OH, Fmoc-Phe-OH, Fmoc-Glu(*t*Bu)-OH, Fmoc-Leu-OH, and Fmoc-Asp(O*t*Bu)-OH, pre-activated with TBTU and NEM, were sequentially coupled to the resin for 1 h. Couplings were repeated to drive the reaction to completion. The above-described mixing, splitting, and coupling steps were repeated until the peptides consisted of eight residues. Details about the amino acids used are shown in Table 1. Finally, after removing the N-terminal Fmoc group, side chain deprotection was performed by treatment with aqueous trifluoroacetic acid ($TFA/H_2O$/triisopropysilane, 95:3:2). The library was washed with 95% aqueous acetic acid solution (3 vol), methanol (3 vol), $CH_2Cl_2$ (3 vol), and water (30 vol).

### Library screening

The library was screened in two separate stages. First, the library resin and the microbeads with target peptide were incubated in separate containers with gentle agitation in water for 2 h. Then, the library resin and target resin were mixed in water and the mixture was incubated for 12 h with gentle stirring (150 r $min^{-1}$), avoiding grinding of the beads. Hits visualized by fluorescence microscopy were manually picked, transferred to a column, and

washed with PBS buffer (50 mM, pH 7.4). In the second stage, the hits were incubated in PBS buffer for 12 h. Hits that remained covered with microbeads were isolated, providing pairs displaying strong recognition.

### Peptide release from single beads and MS analysis

The small $PEGA_{800}$ beads containing the target peptide were removed from the positive library beads with a syringe needle, by using a microscope for visualization. The library beads were transferred to separate wells of a 96-well plate. Release of the peptides was achieved by addition of aqueous TEA solution (10 μL, 5%). The cleavage mixture was allowed to react for 2 h at room temperature. The solutions were transferred to Eppendorf tubes. All beads were washed twice with aqueous acetonitrile solution (20 μL, 70%) for 10 min. Wash fractions were combined with the respective cleavage solutions. The samples were lyophilized twice. Water was added prior to the second lyophilization step to remove the majority of TEA. Samples were analyzed by MALDI-MS as described in the Supporting Information.

### Resynthesis of hit sequences and verification of binding

Following the general SPPS procedure, the positive hits from the library and three consensus sequences were synthesized on 0.1 g $PEGA_{1900}$ resin. Following deprotection, the resin was incubated in water for two hours. During this incubation, the water was exchanged three times. Small amounts of the ligand-containing beads were transferred to a 96-well plate. Binding under the conditions of water, urea solution, high salt, and FBS was investigated both for assembly of microbeads on macrobeads and for dissociation of pre-assembled clusters. The beads were incubated for 0.5 h in water, 0.4 M NaCl aqueous solution, 0.4 M urea solution, and 10% FBS aqueous solution. The density of small, target-containing beads on the larger library beads was used as a qualitative measure of the binding affinity. The recognition was shown to be specific, as unrelated peptides did not show any binding in these assays.

### Synthesis of the fluorescein-labeled target

The amino fluorescein-labeled target was synthesized as follows. HMBA-modified $PEGA_{800}$ resin (0.5 g) was esterified with Fmoc-Gly-OH by using 1-(2-mesitylenesulfonyl)-3-nitro-1*H*-1,2,4-triazole (MSNT) and *N*-methylimidazole. For subsequent couplings, the Fmoc group was removed with 20% piperidine in DMF and the amino acid derivatives were activated as described above with TBTU and NEM. Coupling reaction times were 2 h. The second amino acid was *tert*-butoxycarbonyl (Boc)-protected Dap for attachment of the fluorophore. After coupling of Fmoc-Dap(Boc)-OH (204.6 mg, 0.48 mmol) by using TBTU (148 mg, 0.46 mmol) and NEM (112 μL, 0.64 mmol), the Boc group was removed by 95% TFA (1.5 h) and washed with $CH_2Cl_2$ (6 vol), methanol (6 vol), and $CH_2Cl_2$ (1 vol). The $NH_2$ group was protected with an allyloxycarbonyl (Aloc) group by following the procedure of Cilibrizzi.[31] The resin was washed with cold $CH_2Cl_2$ and allyl chloroformate (Aloc-Cl, 20 μL, 0.2 mmol, 2.0 equiv) in cold $CH_2Cl_2$ was added and reacted for 3 h. Then, the resin was washed with $CH_2Cl_2$ (3 vol), methanol (3 vol), and DMF (3 vol). The other amino acids were coupled by following the general SPPS procedure. Upon completion of the peptide synthesis, the Aloc group was removed according to the procedure by Thieriet.[32] The beads were washed with $CH_2Cl_2$, and $PhSiH_3$ (24 equiv, in $CH_2Cl_2$) was added under $N_2$. The suspension was bubbled with $N_2$. $Pd(PPh_3)_4$ (0.1 equiv, in $CH_2Cl_2$) was added and the reaction was allowed to take place for 15 min under $N_2$ bubbling. The resin was washed and the process was repeated. The reaction was monitored by using the Kaiser test. Following washing with $CH_2Cl_2$ (3 vol), methanol (3 vol), and DMF (3 vol), a solution of amino-fluorescein succinate (134.3 mg, 0.3 mmol, 3.0 equiv) was added and reacted for 2 h. After washing with DMF (3 vol), methanol (3 vol), and $CH_2Cl_2$ (3 vol), the peptide was deprotected by using 20% piperidine and 95% TFA following the general SPPS procedure. The fluorescent-labeled peptide was purified by preparative HPLC and analyzed by HRMS-ESI (*m/z* $[M+2H^+]$: calcd 879.36676, found 879.36649).

### Determination of $K_d$ values

The fluorescently labeled target molecule (20.82 mg, 0.0118 mmol) was dissolved as a 10 mM stock solution. This solution was diluted in Milli-Q purified water to concentrations ranging from 0.001 μM to 100 μM. A single bead was isolated and immobilized in the microchannel of a custom-made flow device (see the Supporting Information). The fluorescence intensity was monitored by using an Olympus IC73 fluorescence microscope. First, the bead labeled with ligand **7** was incubated with the second highest (50 μM) and highest (100 μM) concentrations of target solution. At these two concentrations, the fluorescence intensity of the bead was shown to be constant, hence, it was defined as the fluorescence intensity at the point of saturation ($I_{F\,max}$). Unbound ligand was removed, providing an $I_{F\,max}$ of 5300. The target was completely eluted from the bead in less than 1 min by using aqueous 0.1% TFA. The bead was washed with Milli-Q water and the background intensity in the absence of the fluorophore was determined. The bead was incubated with a flow of 1 nM target solution until the fluorescence intensity was constant. This was followed by flow of increasing target concentrations (0.002, 0.004, 0.008, 0.016, 0.032, 0.064, 0.1, 0.2, 0.4, 0.6, 0.7, 0.8, 1, 2, 3, 5, 10, 15, 20 μM) until constant intensity. Each measurement was performed in triplicate and averaged. The binding constants of ligands **17**, **18**, and **19** were determined by using this same method.

### Recognition measurements by 2D NMR spectroscopy

Solution-state NMR spectroscopy, in particular NOESY spectra, was utilized to investigate the interactions between the two peptides. All NMR spectra were acquired by using standard pulse sequences on a Bruker Avance 800 spectrometer at 298 °C. Both pre-saturation and watergate pulses were used for water suppression and the resulting spectra were compared to ensure that the suppression did not affect the cross-peaks of, for example, the NOESY spectra. The individual spectra were fully assigned from phase-sensitive DQF-COSY, NOESY, and HSQC experiments. Spectra were acquired for the individual peptides and for a mix of the two. The pure target **20** (6.97 mg, 5.88 μmol) was dissolved in 0.5 mL $D_2O$/$H_2O$ (1:9). The pure ligand **7** (5.43 mg, 5.61 μmol) was dissolved in 0.5 mL $D_2O/H_2O$ (1:9) and the pH was adjusted to 5.5, 6.0, or 6.5 by using 1 M NaOH solution. The pH of the target solution was adjusted to 5.5, 6.0, or 6.5 by using NaOH solution. Oxygen was removed by bubbling $N_2$ through the solutions for 15 min. Following acquisition of the individual NMR spectra of the ligand and target peptides, the solutions were mixed. Oxygen was again removed by $N_2$ bubbling. The mixture (0.5 mL) was transferred to an NMR tube and subsequently NOESY and ROESY (rotating frame NOE spectroscopy) spectra were acquired. Acquisition parameters and conditions, as well as tables with observed distances, correlation times, chemical shift differences, and experimental procedures of the NMR titrations are presented in the Supporting Information.

## Acknowledgments

Our thanks go to Frederik Diness, Surendra Vutti, and Theis Brock-Nannestad for technical assistance. This work was supported by the China Scholarship Council (CSC) and Copenhagen University, CECB.

**Keywords:** binding assay · charge–charge interactions · combinatorial selection · molecular recognition · peptides

[1] S. Woo, P. W. K. Rothemund, *Nat. Chem.* **2011**, *3*, 620 – 627.
[2] Y. Levy, J. N. Onuchic, *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 389 – 415.
[3] G. R. Crabtree, N. A. Clipstone, *Annu. Rev. Biochem.* **1994**, *63*, 1045 – 1083.
[4] E. J. Sundberg, R. A. Mariuzza, *Adv. Protein Chem.* **2002**, *61*, 119 – 160.
[5] S. J. Wodak, J. L. Janin, *Adv. Protein Chem.* **2002**, *61*, 9 – 73.
[6] R. E. Babine, S. L. Bender, *Chem. Rev.* **1997**, *97*, 1359 – 1472.
[7] B. W. Sigurskjold, D. R. Bundle, *J. Biol. Chem.* **1992**, *267*, 8371 – 8376.
[8] J. F. Couture, E. Collazo, R. C. Trievel, *Nat. Struct. Mol. Biol.* **2006**, *13*, 698 – 703.
[9] H. Kaur, M. Datt, M. K. Ekka, M. Mittal, A. K. Singh, S. Kumaran, *Biochimie* **2011**, *93*, 175 – 186.
[10] E. Diamandis, T. K. Christopoulos, *Clin. Chem.* **1991**, *37*, 625 – 637.
[11] T. Young, R. Abel, B. Kim, B. J. Berne, R. A. Friesner, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 808 – 813.
[12] X. Cha, K. Ariga, T. Kunitake, *J. Am. Chem. Soc.* **1996**, *118*, 9545 – 9551.
[13] M. Hossain, H.-J. Schneider, *J. Am. Chem. Soc.* **1998**, *120*, 11208 – 11209.
[14] W. J. Cooper, M. L. Waters, *Curr. Opin. Chem. Biol.* **2005**, *9*, 627 – 631.
[15] J. Eichler, *Comb. Chem. High Throughput Screening* **2005**, *8*, 135 – 143.
[16] Y. Ahn, Y. Jang, N. Selvapalam, G. Yun, K. Kim, *Angew. Chem. Int. Ed.* **2013**, *52*, 3140 – 3144; *Angew. Chem.* **2013**, *125*, 3222 – 3226.
[17] I. Hwang, K. Baek, M. Jung, Y. Kim, K. M. Park, D. W. Lee, N. Selvapalam, K. Kim, *J. Am. Chem. Soc.* **2007**, *129*, 4170 – 4171.
[18] M. Lisbjerg, B. E. Nielsen, B. O. Milhoj, S. P. A. Sauer, M. Pittelkow, *Org. Biomol. Chem.* **2015**, *13*, 369 – 373.
[19] T. Haack, M. W. Peczuh, X. Salvatella, J. Sánchez-Quesada, J. de Mendoza, A. D. Hamilton, E. Giralt, *J. Am. Chem. Soc.* **1999**, *121*, 11813 – 11820.
[20] C. Schmuck, L. Geiger, *J. Am. Chem. Soc.* **2004**, *126*, 8898 – 8899.
[21] Y. Zhao, J. Gu, Y. C. Yang, H. Shi, H. Y. Zhu, R. Huang, B. Jing, *Res. Chem. Intermed.* **2009**, *35*, 597 – 605.
[22] M. Monnee, A. Brouwer, R. Liskamp, *QSAR Comb. Sci.* **2004**, *23*, 546 – 559.
[23] K. H. Chang, J. H. Liao, C. T. Chen, B. K. Mehta, P. T. Chou, J. M. Fang, *J. Org. Chem.* **2005**, *70*, 2026 – 2032.
[24] S. Sasaki, M. Takagi, Y. Tanaka, M. Maeda, *Tetrahedron Lett.* **1996**, *37*, 85 – 88.
[25] M. Meldal, B. Q. Wu, F. Diness, R. Michael, G. Hagel, *ChemBioChem* **2011**, *12*, 2463 – 2470.
[26] M. Meldal, S. F. Christensen, *Angew. Chem. Int. Ed.* **2010**, *49*, 3473 – 3476; *Angew. Chem.* **2010**, *122*, 3551 – 3554.
[27] R. Bachor, M. Cydzik, M. Rudowska, A. Kluczyk, P. Stefanowicz, Z. Szewczuk, *Mol. Diversity* **2012**, *16*, 613 – 618.
[28] L. S. Simpson, T. Kodadek, *Tetrahedron Lett.* **2012**, *53*, 2341 – 2344.
[29] F. I. Auzanneau, M. Meldal, K. Bock, *J. Pept. Sci.* **1995**, *1*, 31 – 44.
[30] K. Frische, M. Meldal, O. Werdelin, S. Mouritsen, T. Jensen, L. Galli-Stampino, K. Bock, *J. Pept. Sci.* **1996**, *2*, 212 – 222.
[31] A. Cilibrizzi, A. Isidro-Llobet, N. Mateu, W. R. J. D. Galloway, D. R. Spring, *Synlett* **2012**, 290 – 294.
[32] N. Thieriet, J. Alsina, E. Giralt, F. Guibé, F. Albericio, *Tetrahedron Lett.* **1997**, *38*, 7275 – 7278.

Received: January 18, 2016
Published online on April 13, 2016

**Homomorphosins A-F, novel diketopiperazines**

Casper Hoeck[1], Lene M. Petersen[2], Thomas O. Larsen[2] and Charlotte H. Gotfredsen[1,*]

[1] Department of Chemistry, Kemitorvet B201, Technical University of Denmark, Kgs. Lyngby DK-2800, Denmark; E-Mails: casho@kemi.dtu.dk (C.H.)

[2] Natural Product Chemistry, Department of Systems Biology, Søltofts Plads B221, Technical University of Denmark, Kgs. Lyngby DK-2800, Denmark; E-Mails: lmape@bio.dtu.dk (L.M.P.); : tol@bio.dtu.dk (T.O.L.)

* Author to whom correspondence should be addressed; E-Mail: chg@kemi.dtu.dk.

Keywords: NMR, $^{1}H$, $^{13}C$, *Aspergillus homomorphus*, Aspergilli, natural products, secondary metabolism, diketopiperazine

**Abstract**

Six novel secondary metabolites, homomorphosins A-F, were isolated from the filamentous fungus *Aspergillus homomorphus*. The metabolites share structural features and the majority of compounds were discovered based on a proposed shared biosynthesis origin and the dereplication of proposed precursors. The structures of the compounds were elucidated using UHPLC-DAD-MS and especially NMR spectroscopy. Their respective stereochemistry was elucidated based on NOESY/ROESY data, back calculation of distances and chemical analysis using Marfey's reagent.

**Introduction**

*Aspergilus homomorphus* is a lesser studied *Aspergillus* species residing in section *Nigri* (the black Aspergilli), which includes fungi such as *A. niger* and *A. carbonarius*.[1] Black aspergilli are of industrial use and known to cause food spoilage.[2] An example of the latter is *A. niger* which is used in large scale production of citric acid, but may also produce ochratoxin A and fumonisin $B_2$.[3,4] *A. homomorphus* was first identified by R. Steiman *et al.* from soil near the Dead Sea in Israel, and belongs in the same clade as *A. aculeatus*. [1,5]

Through the years a wide variety of secondary metabolites (SMs) have been reported from black Aspergilli.[6] The continued search for new chemical entities (NCEs) in the section is thus spurred by the ability of related species to produce quite different SMs.[7,8] *A. homomorphus* have previously been reported to produce secalonic acid D and F as well as dehydrocarolic acid.[1] Here

the SM chemistry of *A. homomorphus* is explored using UHPLC-DAD-MS and NMR spectroscopy in the continued search for both NCEs and the elucidation of their biosynthetic origin.

Most of the reported structures share features and were identified through a proposed shared biosynthetic relationship, which was used to guide the identification and isolation of novel precursor compounds. This was proposed without any knowledge from genetics, as the genome of *A. homomorphus* has not yet been explored.

**Results and discussions**

The structures of the metabolites isolated and characterized from *A. homomorphus* are shown in Figure 1. Most of these are presumably altered non-ribosomal peptides (NRPs) with attached isoprene units. They are thereby related to other natural compounds such as okaramines, which are produced by other black aspergilli, e.g. *A. aculeatus* and *A. indologenus*, as well as *Penicillium simplicissimum*.[9–14] All relevant MS, UV and NMR data are found in the Supporting Information.

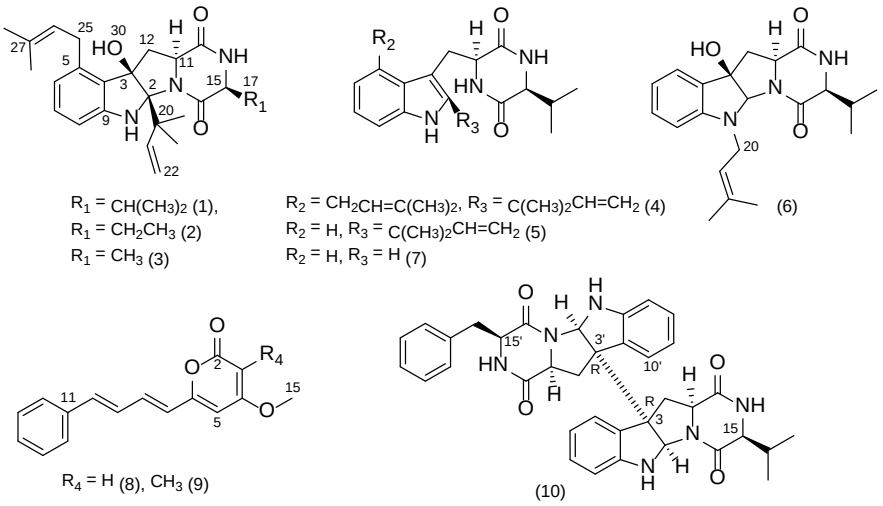



**Figure 1.** Novel compounds from *A. homomorphus* (**1-6**), compounds previously reported as "unnatural" (**8**,**9**)[15,16] and rediscovered compounds (**7**,**10**)[17,18]. Compound **7** is included due to its likely biosynthetic relationship with **1**, **4**, **5** and **6**. Data of **8-10** may be found in the Structural data section. Also secalonic acid D and F were reconfirmed to be produced via UHPLC-DAD-MS in accordance with previous publications.[1]

*Novel metabolites*

The common name of the novel compounds, homomorphosin, was chosen due to their fungal origin. The NMR data for the homomorphosins are given in Table 1.

**Table 1.** $^{1}$H NMR (500 or 800 MHz) and $^{13}$C NMR data (125 or 200 MHz) for **1-6** (in DMSO-$d_6$. δ in ppm and *J* in Hz)

| # | **1** | | **2** | | **3** | | **4** | | **5** | | **6** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\delta_H$ (mult, J) | $\delta_C$ | $\delta_H$ (mult, J) | $\delta_C$ | $\delta_H$ (mult, J) | $\delta_C$ | $\delta_H$ (mult, J) | $\delta_C$ | $\delta_H$ (mult, J) | $\delta_C$ | $\delta_H$ (mult, J) | $\delta_C$ |
| 1 | 6.72 (s) | | 6.69 (1,s) | | 6.66 (s) | | 10.59 (s) | | 10.51 (s) | | | |
| 2 | | 90.8 | | 90.7 | | 90.3 | | 141.2 | | 140.9 | 5.31 (s) | 88.0 |
| 5 | | 139.3 | | 139.3 | | 139.3 | | 132.7 | 7.42 (d,7.8) | 117.8 | 7.17 (d,7.2) | 122.3 |
| 6 | 6.47 (d,7.6) | 119.3 | 6.47 (d,7.4) | 119.5 | 6.47 (d,7.4) | 119.6 | 6.69 (d,7.1) | 119.2 | 6.92 (t,7.1) | 118.2 | 6.72 (t,7.4) | 118.5 |
| 7 | 6.99 (t,7.7) | 129.4 | 6.99 (t,7.7) | 129.3 | 6.99 (t,7.7) | 129.5 | 6.92 (t,7.6) | 120.4 | 7.00 (t,7.1) | 120.3 | 7.11 (t,7.5) | 129.7 |
| 8 | 6.56 (d,7.8) | 107.8 | 6.57 (d,7.6) | 108.1 | 6.57 (d,7.4) | 108.2 | 7.21 (d,7.9) | 108.9 | 7.30 (d,8.0) | 110.6 | 6.45 (d,7.9) | 108.7 |
| 10 | | | | | | | 6.84 (s) | | 7.49 (1,d,3.0) | | | |
| 11 | 3.66 (dd,11.5,7.3) | 57.0 | 3.68 (dd,11.5,7.3) | 57.3 | 3.72 (dd,11.5,7.5) | 57.4 | 4.03 (d,10.4) | 55.9 | 3.99 (m) | 55.4 | 4.57 (dd,11.6,6.3) | 59.2 |
| 12a | 2.54 (m) | | 2.53 (m) | | 2.53 (m) | | 3.13 (dd,14.8,10.7) | | 2.99 (dd,14.5,9.2) | | 1.78 (m) | |
| 12b | 2.62 (dd,13.0,7.4) | | 2.62 (m) | | 2.63 (dd,11.5,7.4) | | 3.57a (m) | | 3.41 (dd,14.5,4.3) | | 2.40b (m) | |
| 14 | 7.95 (s) | | 8.05 (s) | | 8.10 (s) | | 8.12 (s) | | 8.18 (d,2.7) | | 7.98 (s) | |
| 15 | 3.77 (d,1.9) | 60.4 | 3.84 (t,4.5) | 56.1 | 3.89 (q,6.8) | 51.3 | 3.57b (m) | 59.2 | 3.48 (m) | 59.6 | 4.02 (m) | 60.4 |
| 17a | 2.32 (septd,7.1,2.0) | | 1.68 (m) | | 1.16 (m) | | 2.13 (m) | | 1.98 (dd,12.4,6.8) | | 2.40a (m) | |
| 17b | | | 1.71 (m) | | | | | | | | | |
| 18 | 0.86 (d,6.8) | 15.8 | 0.85 (t,7.3) | 8.7 | | | 0.91 (d,6.7) | 17.4 | 0.89 (d,6.8) | 17.4 | 0.87 (d,6.8) | 16.5 |
| 19 | 0.98 (d,7.2) | 18.1 | | | | | 0.99 (d,7.0) | 18.5 | 0.96 (d,6.9) | 18.4 | 1.02 (d,7.2) | 18.5 |
| 20a | | | | | | | | | | | 3.93 (m) | |
| 20b | | | | | | | | | | | 4.00 (m) | |
| 21 | 6.27 (dd,17.6,10.8) | 145.2 | 6.27 (dd,17.6,10.9) | 145.4 | 6.26 (dd,17.6,10.8) | 145.6 | 6.16 (dd,17.5,10.5) | 146.7 | 6.15 (dd,17.5,10.5) | 146.3 | 5.08 (t,5.8) | 119.9 |
| 22a | 4.85 (m) | | 4.84 (d,11.1) | | 4.85 (dd,12.8,1.7) | | 5.03 (d,11.0) | | 5.00 (m) | | | |
| 22b | 4.85 (m) | | 4.87 (d,17.6) | | 4.88 (dd,17.9,1.7) | | 5.04 (d,17.5) | | 5.03 (m) | | | |
| 23 | 1.17 (s) | 24.2 | 1.17 (s) | 24.2 | 1.17 (s) | 24.2 | 1.49 (s) | 27.3 | 1.47 (s) | 27.5 | 1.60 (s) | 26.0 |
| 24 | 1.20 (s) | 24.5 | 1.20 (s) | 24.6 | 1.19 (s) | 24.6 | 1.52 (s) | 27.8 | 1.48 (s) | 27.5 | 1.67 (s) | 18.4 |
| 25a | 3.33 (m) | | 3.34 (m) | | 3.34 (m) | | 3.67 (m) | | | | | |
| 25b | 3.47 (dd,15.6,7.3) | | 3.46 (dd,15.7,6.9) | | 3.47 (dd,15.7,7.5) | | 3.75 (m) | | | | | |
| 26 | 5.22 (t,6.9) | 123.0 | 5.23 (1,td,6.7,1.4) | 123.1 | 5.23 (t,6.9) | 123.2 | 5.21 (t,6.7) | 123.9 | | | | |
| 28 | 1.70 (s) | 17.7 | 1.70 (s) | 25.2 | 1.71 (s) | 25.3 | 1.68a (s) | 17.7 | | | | |
| 29 | 1.70 (s) | 25.2 | 1.69 (s) | 17.6 | 1.70 (s) | 17.6 | 1.68b (s) | 25.5 | | | | |
| 30 | 5.70 (s) | | 5.69 (s) | | 5.69 (s) | | | | | | | |

Homomorphosin A (**1**) was isolated as a white solid. The $[M+H]^+$ of **1**, *m/z* 438.2754, corresponds to a constituent formula of $C_{26}H_{36}N_3O_3$ (theoretical *m/z* 438.2751). The 35 protons found in the 1D $^1H$ spectrum were divided among six methyl groups at $\delta_H$ 0.86 (3H, d, *J*=6.8, H-18), 0.98 (3H, d, *J*=7.2, $H_3$-19), 1.17 (3H, s, $H_3$-23), 1.20 (3H, s, $H_3$-24) and 1.70 (6H, s, $H_3$-28/29), three aromatic protons at $\delta_H$ 6.47 (1H, d, *J*=7.6, H-6), 6.56 (1H, d, *J*=7.8, H-8) and 6.99 (1H, t, *J*=7.7, H-7), four olefinic protons at $\delta_H$ 4.85 (2H, m, $H_2$-22), 5.22 (1H, t, *J*=6.9, H-26) and 6.27 (1H, dd, *J*=17.6, 10.8, H-21), two methines with neighboring heteroatoms at $\delta_H$ 3.66 (1H, dd, *J*=11.5, 7.3, H-11) and 3.77 (1H, d, *J*=1.9, H-15), two diastereotopic methylenes at $\delta_H$ 2.54 (1H, m, H-12a), 2.62 (1H, dd, *J*=13.0, 7.4, H-12b), 3.33 (1H, m, H-25a) and 3.47 (1H, dd, *J*=15.6,7.3, H-25b), a methine at $\delta_H$ 2.32 (1H, septd, *J*=7.1, 2.0, H-17) along with exchangeable protons at $\delta_H$ 5.70 (1H, s, H-30), 6.72 (1H, s, H-1) and 7.95 (1H, s, H-14). All protons could be coupled to their respective carbon resonances (as in Table 2) through gHSQC correlations. The diastereotopic protons and the resonance at 1.70 ppm could thus be elucidated. Five spin systems were established from the DQF-COSY experiment; an aromatic, an olefinic and three aliphatic, and the spin systems and quaternary carbons were joined through gHMBC correlations as illustrated in Figure 2a.

The relative stereochemistry of **1** is suggested from a network of NOEs obtained from NOESY and ROESY experiments with the key NOE being H-11/H-15, which established that the amino acids share handedness. NOE correlations between H-11/H-12b, H-12a/H-30, H-12a/$H_3$-23/24 and H-21/H-30 were all used to solve the remaining relative stereochemistry. The coupling constants of H-11 and H-12a/b fit to the proposed stereochemistry of the methylene group with a smaller $^3J$-coupling originating from H-11/H-12b in comparison to H-11/H-12a, which shows a large $^3J$-coupling constant. Figure 2b includes the important NOEs, while 2c shows the 3D structure with the best fit to the correlating distances obtained by the isolated spin pair approximations (ISPA) and the observed coupling constants. The absolute configuration was solved by hydrolyses and reaction with Marfey's reagent which suggested an L-configuration of valine.[19] The $^3J$-coupling constant between H-15 and H-17 is 2.0 Hz which suggests a restriction in the rotation of the isopropyl group attached to C-1.

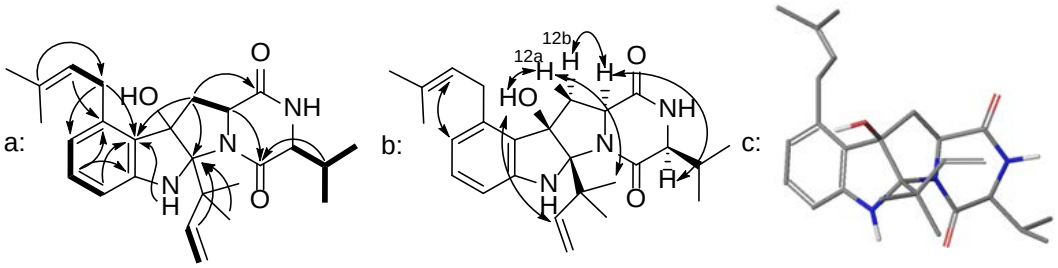



**Figure 2.** a: COSY spin systems (**—**) and chosen HMBC correlations (→) of **1**. b: Chosen NOEs (→) of **1**. c: The 3D structure of **1** which fit the NOESY data the best.

*Biosynthesis of the homomorphosins*

The isolation of **1** led to a hypothetical biosynthesis as proposed in Figure 3, which in turn led to the discovery of more NRPs. This methodology of discovering new metabolites was used, as **1** was by far the main metabolite produced on the media used. The biosynthesis was proposed utilizing known reactions from organic and biochemical synthesis, starting from L-valine and L-tryptophan (not shown in figure). The mechanisms needed to obtain the compounds are two *C*-prenylations, an epoxidation and subsequent an intramolecular ring-closure. Both mechanisms have been described previously or suggested in the biosynthesis of other natural products. [20–29]

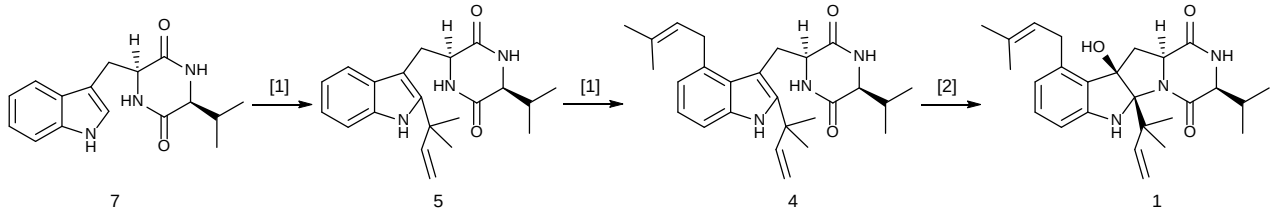



**Figure 3.** The suggested biosynthesis of **1**. Compound **7** is proposed to be made by the cyclization of L-valine and L-tryptophan (not shown). [1] *C*-prenylation, [2] epoxidation/ring closure.

A dereplication, utilizing UHPLC-DAD-MS (chromatograms in Figure 4), was then conducted in order to identify the proposed stable intermediates, which were all isolated and structure elucidated.

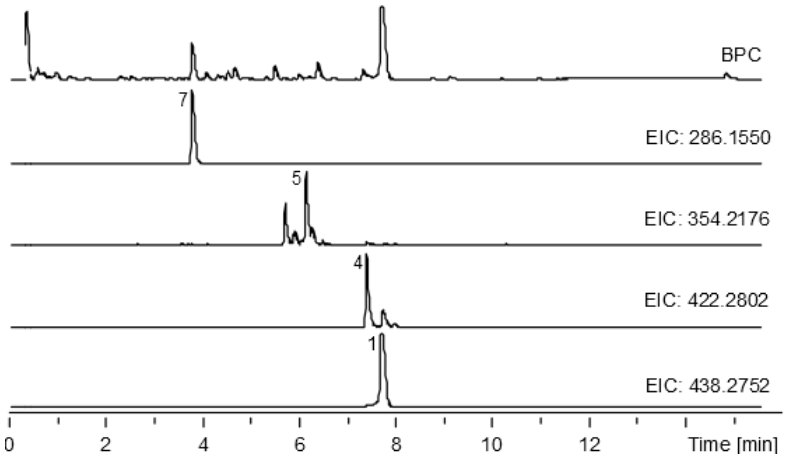



**Figure 4.** Base peak chromatogram (BPC) of extract of *A. homomorphus* (top) and extracted ion chromatograms (EICs) of the *m/z* values for the suggested intermediates with a mass tolerance of ±0.005 Da in the biosynthesis of **1**. Chromatograms are scaled.

The novel diketopiperazines homomorphosins D (**4**) and E (**5**) as well as L-valyl-L-tryptophan anhydride (**7**) were isolated as white solids. The NMR data resembles that of **1** where structural features are shared, and the data of the rediscovered compound **7** corresponds to literature data.[17] The NMR data and $UV_{max}$ suggested that these compounds contain an indole ring which differentiates them from **1**.[30] It is speculated that the compounds **4**, **5** and **7** share a biosynthetic pathway though the genome sequence has yet to be explored.

Due to the successful identification of precursors for **1**, a new round of dereplication to identify structural analogues of **1** was conducted, with tryptophan coupled to other amino acids, which revealed compounds **2** and **3**.

Homomorphosins B (**2**) and C (**3**) were both white solids. The NMR data corresponds to that of **1**, except that valine is exchanged with alanine (**3**) or the "unnatural" amino acid 2-aminobutyrate (**2**). The latter amino acid is not commonly found in natural diketopiperazines, although some have been published, which were functionalized at the α-position.[31] The $^{3}J$-coupling of H-15 and H-17 increases as the size of the substituent attached to C-15 decreases, which may support the suggested rotational restriction for **1**.

Homomorphosin F (**6**) was isolated as a white solid (data found in supporting info). The compound differs from the other homomophosins, and compounds isolated from *A. homomorphus* in general, as it is *N*-prenylated. As for **1**, **2** and **3** the $UV_{max}$ of **6** showed a reduction of the indole ring. The alcohol $^{1}H$ resonance did not appear in the NMR spectra, probably due to a high water content of

the sample. The proposed structure was mainly elucidated based on the constituent formula obtained from HRMS, the chemical shift of C-3 and the UV spectrum. The respective UV-spectra are compared in Supporting Information.

The absolute stereochemistry of **2**-**6** is believed to be the same as that of **1** due to a similar relative stereochemistry in between the compounds, and the possible involvement of **4** and **5** in the biosynthesis of **1**. The absolute stereochemistry of **7** was established by Marfey's chemical analysis to be that of **1**. No NOE between H-11 and H-15 was found in the NOESY/ROESY spectra of **3**, which might suggest another relative configuration of the amino acids. This could also be due to the sparse amount of compound, as the sign of the optical rotation is of the same sign as that of the remaining homomorphosins. A DP4 analysis, as suggested by Smith and Goodman,[32] of experimental and calculated shifts from optimized structures also points towards the L,L compound with a probability of 97.7 %. Coupled to the sign of the optical rotation the stereochemistry is deemed equal to that of **1**.

**Conclusion**

A large scale cultivation of *A. homomorphus* led to the discovery of six novel NRP based metabolites with a diketopiperazines skeleton. They were all characterized using UHPLC-DAD-MS and NMR spectroscopy. Based on NOESY and ROESY NMR data, *J*-coupling constants and structural simulations the relative stereochemistry and a possible 3D structure for the compounds have been proposed. For compound **1** the absolute stereochemistry was obtained using Marfey's chemical analysis. It is speculated that all the new compounds have a shared biosynthetic origin. The proposed biosynthetic pathway was used to identify new metabolites and have gained credibility by the identification of the proposed structures by isolation of these and elucidation of their structures.

## Experimental

*Fungal growth, extraction and the isolation of compounds.*

*A. homomorphus* (IBT 21893) was inoculated on different media to determine the preferred growth media and time in order to get as diverse a chemical profile as possible. The used media were yeast sucrose agar (YES), Czapek yeast extract agar (CYA), malt extract agar (MEA), oatmeal agar (OAT) and creatine sucrose agar (CREA) for 2, 5, 7 and 9 days.[33] The major metabolites from most growth media were homomorphosin A (**1**) and 3-methyl-4-methoxy-6-(4-phenyl-1,3-butadienyl)-2H-pyran-2-one (**9**) which could originate from polyketide and Shikimi pathways.[34] The chosen media and time were YES for 7 days.

*A. homomorphus* was inoculated as three-point stabs on agar plates and incubated in the dark at 30 °C for 7 days. The used media were YES for the NRPs (200 plates) and MEA for the other isolated metabolites (50 plates). The fungi were harvested and extracted twice with EtOAc (1 % FA) (once overnight), m(YES) = 7.8 g, m(MEA) = 217.1 mg. The combined extract was dissolved in 9:1 MeOH:$H_2O$ and washed with heptane. Water was added to give 1:1 MeOH/$H_2O$ and compounds were extracted with DCM followed by concentration *in vacuo*, m(YES) = 4.90 g, m(MEA) = 36.3 mg. The DCM phase (YES) was absorbed onto reverse phase column material and fractioned by an Isolera One flash chromatography system by Biotage with auto-fractionation and a 120 g Biotage Snap KP-C18-HS column. The column volume used 132 mL and fractions of 115 mL (max) were collected. The gradient ranged from 30 to 100 % MeCN/$H_2O$ with a flow rate of 40 mL/min over 65 min.

The fractions (see table below) which showed possible targets after UHPLC-DAD-MS analysis were further purified on a semi-preparative HPLC; the Waters 600 Controller coupled to a Waters 996 Photodiode Array Detector with a flow rate of 5 mL/min. The columns used were Luna II C18, 5 μm, 250 × 10 mm by Phenomenex and Gemini C6-Phenyl, 5 μm, 250 × 10 mm by Phenomenex.

**Table 2.** Purification of compounds.

| | Isolera fraction #/%MeCN:$H_2O$ | m [mg] | Waters run %MeCN:$H_2O$/Time [min] | Waters Time [min]/ %MeCN:$H_2O$ | m [mg] |
|---|---|---|---|---|---|
| **1** | 15/70-74 | 180.7 | - | - | 180.7 |
| **2** | 13/64-68 | 72.0 | Iso 65/15 | 7.3/65 | 2.7 |
| **3** | 11/56-60 | 14.6 | Grad 40-70/20 | 14.5/63 | 0.9 |
| **4** | 13/64-68 | 72.0 | Iso 65/15 | 7.7/65 | 3.6 |
| **5** | 8/45-49 | 36.9 | Iso 45/15 | 7.5/45 | 1.7 |
| **6** | 4/30-34 | 72.6 | Grad 30-40/15 | 5.3/34 | 8.1 |
| **7** | 10/52-56 | 27.0 | Iso 40/20 | 15.5/40 | 1.9 |
| **8** | - | 36.3 | Iso 55/20 | 13.4/55 | 1.7 |
| **9** | - | 36.3 | Iso 55/20 | 15.8/55 | 2.5 |
| **10** | 8/45-49 | 36.9 | Iso 45/15 | 6.35/45 | 3.5 |

*NMR*

All NMR spectra were acquired using standard pulse sequences on a Unity Inova 500 by Varian or on a Bruker Avance 800 spectrometer of the Danish Instrument Center for NMR Spectroscopy of Biological Macromolecules at 298 °C.[35–40] The deuterated solvent used for all compounds was DMSO-$d_6$ which was used as internal standard ($\delta_H$ = 2.49 ppm, $\delta_C$ = 39.5 ppm). Typical acquisition conditions were as follows (exemplified by the conditions for homomorphosin A at 500 MHz). 1D $^1$H: Spectrometer frequency (sf) = 499.87 MHz, spectral width (sw) = 5000.0 Hz, Fourier transform size (si) = 20480, acquisition time (aq) = 2.05 s, 90 ° pulse width (pw) = 7.0 µs, relaxation delay (rd) = 2 s, number of scans (ns) = 32, number of dummy scans (ds) = 8. gHSQC: sf = 499.87 MHz, sw($^1$H) = 5000.0 Hz, sw($^{13}$C) = 30007.7 Hz, si = 2048, ni = 512, aq = 0.20 s, pw = 6.9 µs, rd = 1.5 s, ns = 8, ds = 64. DQF-COSY: sf = 499.87 MHz, sw = 5000.0 Hz, si = 2048, ni = 1024, aq = 0.20 s, pw = 6.9 µs, rd = 1.5 s, ns = 8, ds = 16. gHMBC: sf = 499.87 MHz, sw($^1$H) = 5000.0 Hz, sw($^{13}$C) = 34999.7 Hz, si = 2048, ni = 512, aq = 0.20 s, pw = 6.9 µs, rd = 1.5 s, ns = 8, ds = 64. NOESY/ROESY: sf = 499.87 MHz, sw = 5000.0 Hz, si = 2048, ni = 512, aq = 0.20 s, pw = 7 µs, rd = 2 s, ns = 8, ds = 16. All spectra were processed as given, phase corrected and baseline corrected by a 5 degree polynomial function. 1D: Number of points (np) = 65k, windows function (wdw) = exponential, line broadening (lb) = 0.3. 2D: np(F1) = 1k, np(F2) = 4k, wdw(F1,F2) = sine squared, sine bell shift (sbs) (F1,F2) = 2. Reported *J*-coupling constants are reported as extracted from the spectra and may differ due to uncertainties.

*Structural parameters from NMR*

To check the stereochemical assignment of homomorphosin A the isolated spin-pair approximation (ISPA) method, with increased linear range as suggested by Macura *et al.*, was applied.[41] The applied mixing time was 150 ms which was found to be within the linear range (50, 100, 150, 200 ms was used for buildup curve). The volumes from the peaks in the ROESY spectra were extracted by integration in Topspin3.1. For the calculation of *J*-couplings from dihedral angles the Karplus equation for amino acids was used.[42,43] Experimental couplings constants were measured in 1D $^{1}$H or DQF-COSY spectra and averaged if necessary. Structures were generated using Macromodel in the Maestro (Schrödinger Release 2013-2: Maestro, version 9.5, Schrödinger, LLC, New York, NY, 2013). A conformational search was conducted in implicit DMSO, treated as a constant dielectric constant of 46. The mixed mode option was used with 10,000 steps and PRCG as minimization method using the force field MMFFs. Structures within 25 kJ of the found minimum was considered. From the conformational search 364 conformations was found. The structure which fitted the data the best was found by comparing back-calculated distances and *J*-couplings with the found values.

Averaging of NMR parameters was done iteratively over 10,000 steps were each step included the structure which would give the best average fit to the data. NOE to *J*-couplings were weighted 10:1. Distances was averaged according to $r_{ave} = \langle r_i^{-6} \rangle^{1/6}$. The script was written in-house in Matlab (R2013b, 8.2.0.701, The MathWorks, Inc., Natick, MA).

*UHPLC-DAD-MS*

UHPLC-DAD-MS data was acquired by a Maxis 3G UHR-QTOF-MS by Bruker Daltonics (Bremen, Germany) with an electrospray ionization (ESI) source combined with an UltiMate 3000 Rapid Separation LC by system Thermo Scientific Dionex (Sunnyvale, CA, USA). The column used was a Kinetex C18, 100x2.1 mm, 2.6 μm, by Phenomenex (Torrance, CA, USA) maintained at 40 °C. Solvents were acetonitrile (ACN) and $H_2O$ (buffered with 20 mM FA) and a linear gradient from 10 to 100 % (v/v) ACN over 10 min with a flow rate was 0.4 mL/min was used. MS was run in positive mode (ESI+) with a mass range of m/z 100–1000. Sodium formate was used as internal standard for mass calibration. UV/VIS spectra from the DAD were collected at wavelengths from 200-700 nm.

*Marfey's reagent*

Stereoisometry of the amino acid valine was elucidated using Marfey's method.[19] 100 µg of **1** and **7** was hydrolyzed with 200 µL of 6 M HCl at 110 °C for 20 h. To the hydrolysis product and vials with 2.5 µmol of standard D- and L-amino acids were added 50 µL of water, 20 µL of 1 M $NaHCO_3$ solution, and 100 µL of 1 % 1-fluoro-2-4-dinitrophenyl-5-L-alanine amide (FDAA) in acetone and left at 40 °C for 1 h to react. The reaction mixtures were removed from the heat and neutralized with 10 µL of 2 M HCl, and the solutions were diluted with 820 µL of MeOH. See supporting information for more details.

*Structural data*

**Homomorphosin A (1)**: HRMS: *m/z* = 438.2754. $[M+H]^+$, calculated for $[C_{26}H_{35}N_3O_3+H]^+$: *m/z* = 438.2751. Adducts: 460 $[M+Na]^+$. $[\alpha]_{589.3}^{20\,C}$ = -1.37°

**Homomorphosin B (2)**: HRMS: *m/z* = 422.2803. $[M+H]^+$, calculated for $[C_{26}H_{35}N_3O_2+H]^+$: *m/z* = 422.2802. Adducts: 444 $[M+Na]^+$. $[\alpha]_{589.3}^{20\,C}$ = -0.14°

**Homomorphosin C (3)**: HRMS: *m/z* = 354.2181. $[M+H]^+$, calculated for $[C_{21}H_{27}N_3O_2+H]^+$: *m/z* = 354.2176. Adducts: 376 $[M+Na]^+$, 707 $[2M+H]^+$, 729 $[2M+Na]^+$. $[\alpha]_{589.3}^{20\,C}$ = -0.15°

**Homomorphosin D (4)**: HRMS: *m/z* = 424.2599. $[M+H]^+$, calculated for $[C_{25}H_{33}N_3O_3+H]^+$: *m/z* = 424.2595. Adducts: 446 $[M+Na]^+$, 869 $[2M+Na]^+$. $[\alpha]_{589.3}^{20\,C}$ = -0.55°

**Homomorphosin E (5)**: HRMS: *m/z* = 410.2446. $[M+H]^+$, calculated for $[C_{24}H_{31}N_3O_3+H]^+$: *m/z* = 410.2438. Adducts: 432 $[M+Na]^+$, 841 $[2M+Na]^+$. $[\alpha]_{589.3}^{20\,C}$ = -0.66°

**Homomorphosin F (6)**: HRMS: *m/z* = 370.2128. $[M+H]^+$, calculated for $[C_{20}H_{19}N_3O_2+H]^+$: *m/z* = 370.2125. Adducts: none. $[\alpha]_{589.3}^{20\,C}$ = -0.01°

**L-valyl-L-tryptophan anhydride (7)**[17]: HRMS: *m/z* = 286.1548. $[M+H]^+$, calculated for $[C_{16}H_{19}N_3O_2+H]^+$: *m/z* = 286.1550. Adducts: 308 $[M+Na]^+$, 571 $[2M+H]^+$, 593 $[2M+Na]^+$. $[\alpha]_{589.3}^{20\,C}$ = -0.14°

$^1$H NMR (499.9 MHz, DMSO-$d_6$, 25 °C, 2.49 ppm): δ 10.82 (s, 1H, H-1), 7.95 (s, 1H, H-14), 7.85 (s, 1H, H-10), 7.58 (d, *J* = 7.8 Hz, 1H, H-5), 7.27 (d, *J* = 8.1 Hz, 1H, H-8), 7.06 (d, *J* = 2.1 Hz, 1H, H-3), 7.00 (t, *J* = 7.5 Hz, 1H, H-7), 6.92 (t, *J* = 7.4 Hz, 1H, H-6), 4.12 (t, *J* = 4.4 Hz, 1H, H-11), 3.47 (dd, *J* = 3.6, 2.1, 1.6 Hz, 1H, H-15), 3.19 (dd, *J* = 14.4, 5.0 Hz, 1H, H-12b), 3.06 (dd, *J* = 14.4,

4.5 Hz, 1H, H-12a), 1.63 (sept-d, *J* = 6.9, 4.0 Hz, 1H, H-17), 0.59 (d, *J* = 7.0 Hz, 1H, H-18), 0.17 (d, *J* = 6.8 Hz, 1H, H-19)

$^{13}$C NMR (125.7 MHz, DMSO-$d_6$, 25 °C, 39.5 ppm): δ 166.8 (C-13), 166.2 (C-16), 135.5 (C-9), 127.8 (C-4), 124.2 (C-2), 120.4 (C-7), 118.6 (C-5), 118.0 (C-6), 110.8 (C-8), 108.5 (C-3), 59.0 (C-15), 54.9 (C-11), 30.7 (C-17), 28.4 (C-12), 18.0 (C-18), 15.8 (C-19)

**4-Methoxy-6-(4-phenyl-1,3-butadienyl)-2*H*-pyran-2-one (8)**[15]: HRMS: *m/z* = 255.1018. $[M+H]^+$, calculated for $[C_{16}H_{14}O_3+H]^+$: *m/z* = 255.1016. Adducts: 277 $[M+Na]^+$, 531 $[2M+Na]^+$.

$^1$H NMR (499.9 MHz, DMSO-$d_6$, 25 °C, 2.49 ppm): δ 7.55 (d, *J* = 7.4 Hz, 2H, H-12), 7.38 (t, *J* = 7.5 Hz, 2H, H-13), 7.30 (t, *J* = 7.3 Hz, 1H, H-14), 7.15 (mult, 1H, H-8), 7.10 (mult, 1H, H-9), 7.01 (d, *J* = 14.8 Hz, 1H, H-10), 6.44 (d, *J* = 14.6 Hz, 1H, H-7), 6.31 (d, *J* = 2.1 Hz, 1H, H-5), 5.63 (d, *J* = 2.1 Hz, 1H, H-3), 3.83 (s, 3H, H-15)

$^{13}$C NMR (125.7 MHz, DMSO-$d_6$, 25 °C, 39.5 ppm): 170.7 (C-4), 162.3 (C-2), 157.7 (C-6), 137.4 (C-10), 136.1 (C-11), 134.0 (C-8), 128.5 (C-13), 128.3 (C-14), 127.2 (C-9), 126.6 (C-13), 122.5 (C-7), 100.8 (C-5), 88.3 (C-3), 56.1 (C-15)

**3-Methyl-4-Methoxy-6-(4-phenyl-1,3-butadienyl)-2*H*-pyran-2-one (9)**[16]: HRMS: *m/z* = 269.1171. $[M+H]^+$, calculated for $[C_{17}H_{16}O_3+H]^+$: *m/z* = 269.1172. Adducts: 291 $[M+Na]^+$, 559 $[2M+Na]^+$

$^1$H NMR (499.9 MHz, DMSO-$d_6$, 25 °C, 2.49 ppm): δ 7.56 (d, *J* = 7.4 Hz, 2H, H-12), 7.38 (t, *J* = 7.6 Hz, 2H, H-13), 7.30 (t, *J* = 7.3 Hz, 1H, H-14), 7.16 (mult, 1H, H-8), 7.16 (mult, 1H, H-9), 7.00 (d, *J* = 14.6 Hz, 1H, H-10), 6.72 (s, 1H, H-5), 6.46 (d, *J* = 14.5 Hz, 1H, H-7), 3.92 (s, 3H, H-15), 1.82 (s, 3H, H-16)

$^{13}$C NMR (125.7 MHz, DMSO-$d_6$, 25 °C, 39.5 ppm): 165.9 (C-4), 163.3 (C-2), 156.7 (C-6), 137.4 (C-10), 136.4 (C-11), 134.5 (C-8), 128.8 (C-13), 128.5 (C-14), 127.4 (C-9), 126.9 (C-12), 123.0 (C-7), 100.8 (C-3), 96.9 (C-5), 56.6 (C-15), 8.8 (C-16)

**Q-20547-E (10)**[18]: HRMS: *m/z* = 645.2821. $[M+H]^+$, calculated for $[C_{37}H_{36}N_6O_5+H]^+$: *m/z* = 645.2820. Adducts: 667 $[M+Na]^+$. $[\alpha]_{589.3}^{20\,C}$ = -0.86°.

$^1$H NMR (499.9 MHz, DMSO-$d_6$, 25 °C, 2.49 ppm): 8.91 (1H, s, formic acid), 8.05 (1H, s, H-14), 7.96 (1H, s, H-14'), 7.73 (1H, m, H-8), 7.54 (1H, m, H-5), 7.27 (1H, m, H-5'), 7.24 (2H, m, H-7, H-19'), 7.21 (1H, m, H-20'), 7.18 (1H, m, H-19'), 7.15 (2H, m, H-20', H-21'), 7.07 (1H, m, H-6),

6.93 (1H, m, H-7'), 6.70 (1H, s, H-1'), 6.50 (1H, m, H-6'), 6.47 (1H, m, H-8'), 5.97 (1H, s, H-2), 5.09 (1H, s, H-2'), 4.35 (1H, t, 4.6, H-15'), 3.94 (1H, s, H-15), 3.80 (1H, m, H-11), 3.78 (1H, m, H-11'), 3.08 (1H, m, H-17b'), 2.97 (1H, m, H-17a'), 2.73 (1H, m, H-12b), 2.67 (1H, m, H-12b'), 2.32 (1H, m, H-12a), 2.25 (1H, m, H-17), 2.03 (1H, t, 12.0, H-12a'), 0.92 (3H, d, 6.0, H-19), 0.73 (3H, d, 6.4, H-18).

$^{13}$C NMR (125.7 MHz, DMSO-$d_6$, 25 °C, 39.5 ppm): 167.8 (2C, C-16, C-13'), 165.8 (C-13), 164.7 (C-16'), 161.9 (formic acid), 150.6 (C-9'), 141.0 (C-9), 137.0 (C-18'), 131.2 (C-4), 130.4 (C-20'), 129.3 (C-21'), 129.1 (C-7), 129.0 (C-7'), 128.0 (C-19'), 126.6 (C-4'), 126.3 (2C, C-19', C-20'), 124.7 (C-5), 124.5 (C-6), 123.9 (C-5'), 117.6 (C-6'), 115.6 (C-8), 108.7 (C-8'), 76.7 (C-2), 76.6 (C-2'), 59.3 (C-15), 58.5 (C-3'), 57.7 (C-3), 57.4 (C-11'), 57.2 (C-11), 55.1 (C-15'), 36.3 (C-12), 35.9 (C-12'), 35.0 (C-17'), 28.7 (C-17), 18.0 (C-19), 16.3 (C-18)

NOTE: Q-20547-E was isolated with as an amide of the secondary amine at position 5 and formic acid, which was used in the "large scale" extraction procedure (not plugs). By dereplication it was found that this artificial *m/z* peak only appears after extraction and was coupled to the disappearance of the *m/z* of unreacted Q-20547-E.

**Acknowledgement**

We thank the Danish Instrument Center for NMR Spectroscopy of Biological Macromolecules for 800 MHz NMR time.

**References**

[1] R. A. Samson, P. Noonim, M. Meijer, J. Houbraken, J. C. Frisvad, J. Varga, *Stud. Mycol.* **2007**, *59*, 129–145.

[2] M. Hendrickx, H. Beguin, M. Detandt, *Mycoses* **2012**, *55*, 148–155.

[3] E. Schuster, N. Dunn-Coleman, J. C. Frisvad, P. W. M. Van Dijck, *Appl. Microbiol. Biotechnol.* **2002**, *59*, 426–435.

[4] J. C. Frisvad, T. O. Larsen, U. Thrane, M. Meijer, J. Varga, R. a Samson, K. F. Nielsen, *PLoS One* **2011**, *6*, e23496.

[5] R. Steiman, P. Guiraud, L. Sage, F. Seigle-Murandi, *Syst. Appl. Microbiol.* **1995**, *17*, 620–624.

[6] K. F. Nielsen, J. M. Mogensen, M. Johansen, T. O. Larsen, J. C. Frisvad, *Anal. Bioanal. Chem.* **2009**, *395*, 1225–1242.

[7] N. P. Keller, G. Turner, J. W. Bennett, *Nat. Rev. Microbiol.* **2005**, *3*, 937–947.

[8] D. J. Newman, G. M. Cragg, *J. Nat. Prod.* **2012**, *75*, 311–335.

[9] L. M. Petersen, C. Hoeck, J. C. Frisvad, C. H. Gotfredsen, T. O. Larsen, *Molecules* **2014**, *19*, 10898–10921.

[10] Y. Shiono, K. Akiyama, H. Hayashi, *Biosci. Biotechnol. Biochem.* **2000**, *64*, 103–110.

[11] H. Hayashi, Y. Asabu, S. Murao, M. Arai, *Biosci. Biotechnol. Biochem.* **1995**, *59*, 246–250.

[12] S. Murao, H. Hayashi, K. Takiuchi, M. Arai, *Agric. Biol. Chem.* **1988**, *52*, 886–886.

[13] H. Hayashi, K. Takiuchi, S. Murao, M. Arai, *Agric. Biol. Chem.* **1988**, *52*, 2131–2133.

[14] J. Varga, J. C. Frisvad, S. Kocsubé, B. Brankovics, B. Tóth, G. Szigeti, R. A. Samson, *Stud. Mycol.* **2011**, *69*, 1–17.

[15] Z. H. Israili, E. E. Smissman, *J. Org. Chem.* **1976**, *41*, 4070–74.

[16] O. Kayser, W. R. Waters, K. M. Woods, S. J. Upton, J. S. Keithly, A. F. Kiderlen, *Planta Med.* **2001**, *67*, 722–725.

[17] M. S. C. Pedras, K. C. Smith, J. L. Taylor, *Phytochemistry* **1998**, *49*, 1575–1577.

[18] J. L. Popp, L. L. Musza, C. J. Barrow, P. J. Rudewicz, D. R. Houck, *J. Antibiot. (Tokyo).* **1994**, *47*, 411–419.

[19] P. Marfey, *Carlsberg Res. Commun.* **1984**, *49*, 591–596.

[20] M. R. Andersen, J. B. Nielsen, A. Klitgaard, L. M. Petersen, M. Zachariasen, T. J. Hansen, L. H. Blicher, C. H. Gotfredsen, T. O. Larsen, K. F. Nielsen, U. H. Mortensen, *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, E99–107.

[21] U. Metzger, C. Schall, G. Zocher, I. Unsöld, E. Stec, S.-M. Li, L. Heide, T. Stehle, *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 14309–14314.

[22] J. Guo, J. Tan, Y. Wang, H. Wu, C. Zhang, X. Niu, *J. Nat. Prod.* **2011**, *74*, 2278–2281.

[23] L. Zhou, T. Zhu, S. Cai, Q. Gu, D. Li, *Helv. Chim. Acta* **2010**, *93*, 1758–1763.

[24] X. Li, T. Li, H. Dang, B. Wang, *Helv. Chim. Acta* **2008**, *91*, 1888–1893.

[25] P. M. Dewick, *Medicinal Natural Products: A Biosynthetic Approach*, John Wiley & Sons, Ltd, **2009**.

[26] T. Asai, Y.-M. Chung, H. Sakurai, T. Ozeki, F.-R. Chang, Y.-C. Wu, K. Yamashita, Y. Oshima, *Tetrahedron* **2012**, *68*, 5817–5823.

[27] M. A. Hollenhorst, S. B. Bumpus, M. L. Matthews, J. M. Bollinger, N. L. Kelleher, C. T. Walsh, V. Pennsyl, U. States, *J. Am. Chem. Soc.* **2010**, *132*, 15773–15781.

[28] N. Lorenz, J. Olsovská, M. Sulc, P. Tudzynski, *Appl. Environ. Microbiol.* **2010**, *76*, 1822–1830.

[29] T. Udagawa, J. Yuan, D. Panigrahy, Y. H. Chang, J. Shah, R. J. D'Amato, *J. Pharmacol. Exp. Ther.* **2000**, *294*, 421–427.

[30] D. Carić, V. Tomisić, M. Kveder, N. Galić, G. Pifat, V. Magnus, M. Soskić, *Biophys. Chem.* **2004**, *111*, 247–257.

[31] M. Figueroa, T. N. Graf, S. Ayers, A. F. Adcock, D. J. Kroll, J. Yang, S. M. Swanson, U. Munoz-Acuna, E. J. Carcache de Blanco, R. Agrawal, M. C. Wani, B. a Darveaux, C. J. Pearce, N. H. Oberlies, *J. Antibiot. (Tokyo).* **2012**, *65*, 559–564.

[32] S. G. Smith, J. M. Goodman, *J. Am. Chem. Soc.* **2010**, *132*, 12946–59.

[33] R. A. Samson, J. Houbraken, U. Thrane, J. C. Frisvad, B. Andersen, *Food and Indoor Fungi. CBS Laboratory Manual Series 2.*, Utrecht, The Netherlands: CBS KNAW Fungal Biodiversity Centre, **2010**.

[34] I.-K. Lee, B.-S. Yun, *J. Antibiot. (Tokyo).* **2011**, *64*, 349–359.

[35] W. P. Aue, E. Bartholdi, R. R. Ernst, *J. Chem. Phys.* **1976**, *64*, 2229–2246.

[36] U. Piantini, O. W. Sørensen, R. R. Ernst, *J. Am. Chem. Soc* **1982**, *104*, 6800–6801.

[37] G. Bodenhausen, D. J. Ruben, *Chem. Phys. Lett.* **1980**, *69*, 185–189.

[38] G. M. Clore, a M. Gronenborn, a Bax, *J. Magn. Reson.* **1998**, *133*, 216–21.

[39] A. Bax, M. F. Summers, *J. Am. Chem. Soc.* **1986**, *108*, 2093–2094.

[40] S. Macura, R. R. Ernst, *Mol. Phys.* **1980**, *41*, 95–117.

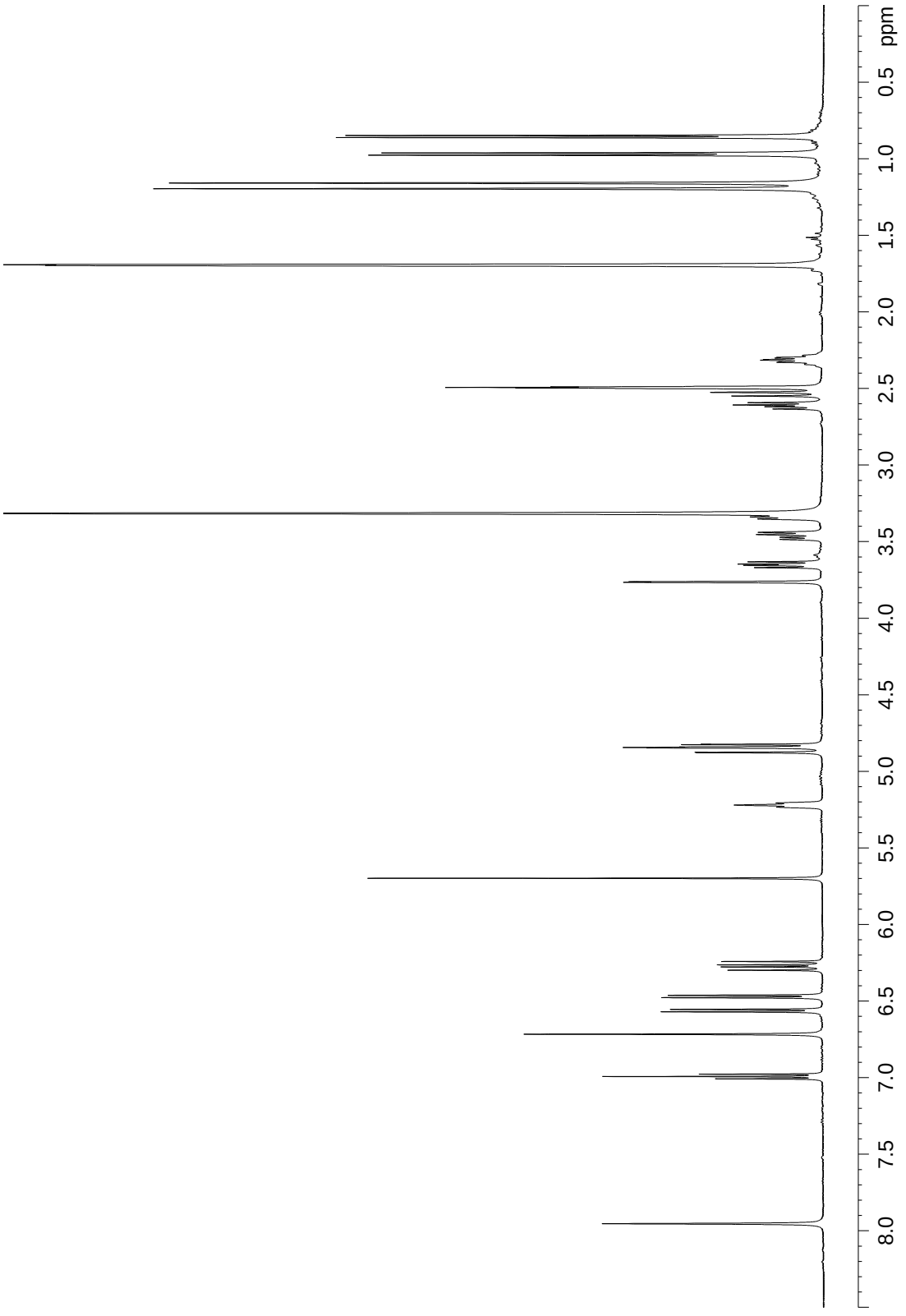
[41] S. Macura, B. T. Farmer II, L. R. Brown, *J. Magn. Reson.* **1986**, *70*, 493–499.

[42] S. Karplus, M. Karplus, *Proc. Natl. Acad. Sci. U. S. A.* **1972**, *69*, 3204–3206.

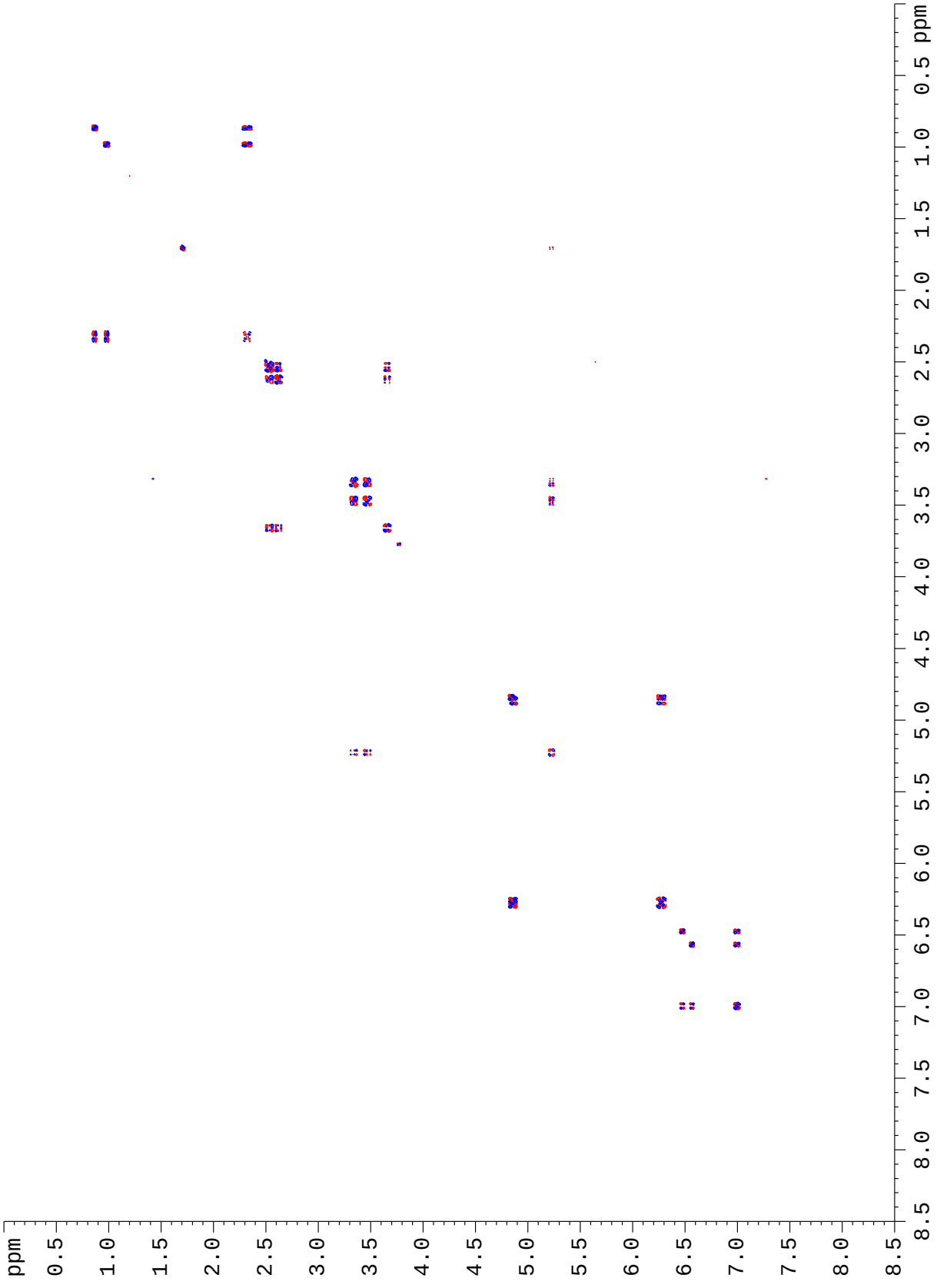
[43] A. Pardi, M. Billeter, K. Wüthrich, *J. Mol. Biol.* **1984**, *180*, 741–751.

**Supporting Information**

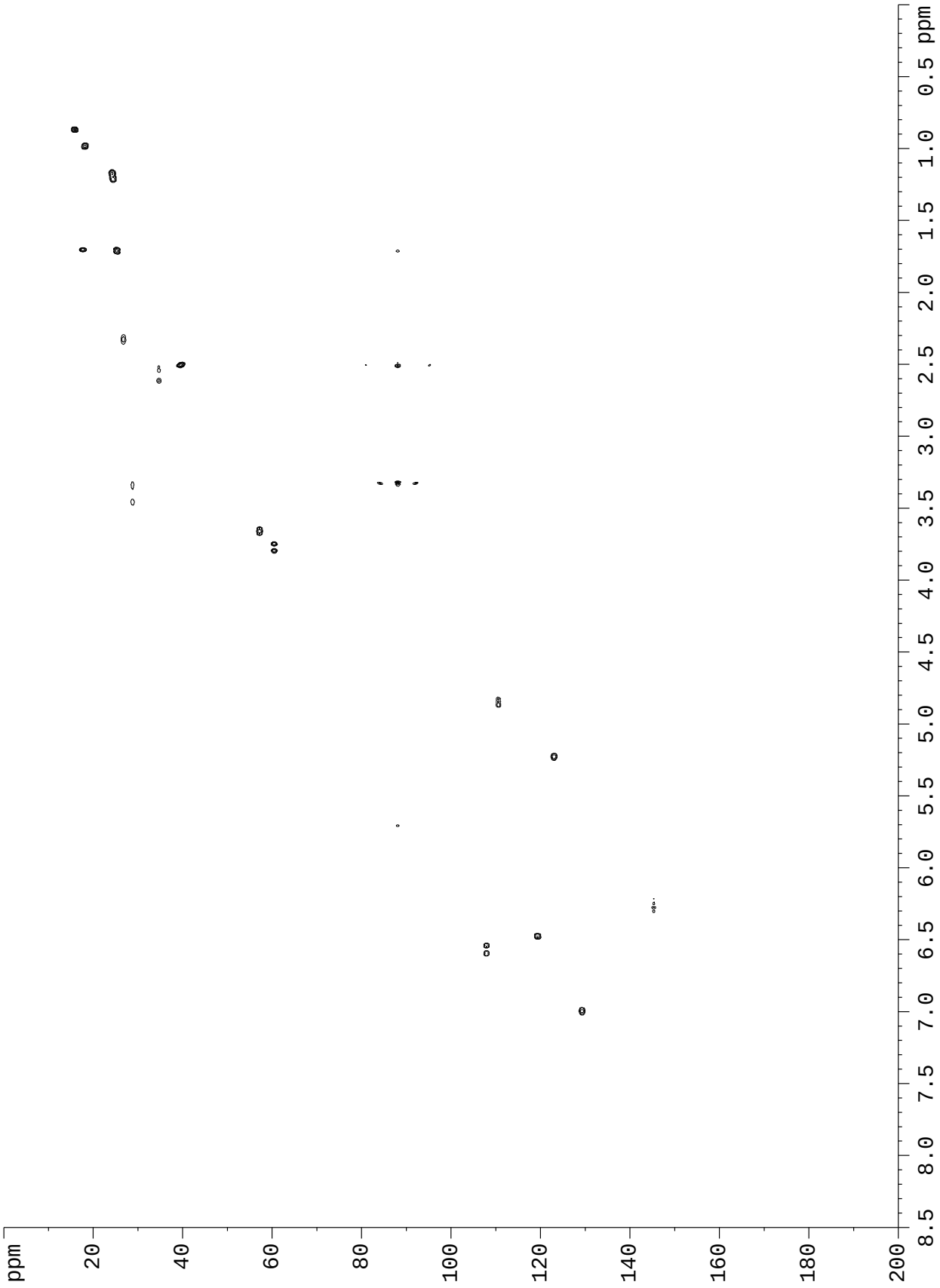
*Spectra of homomorphosin A*



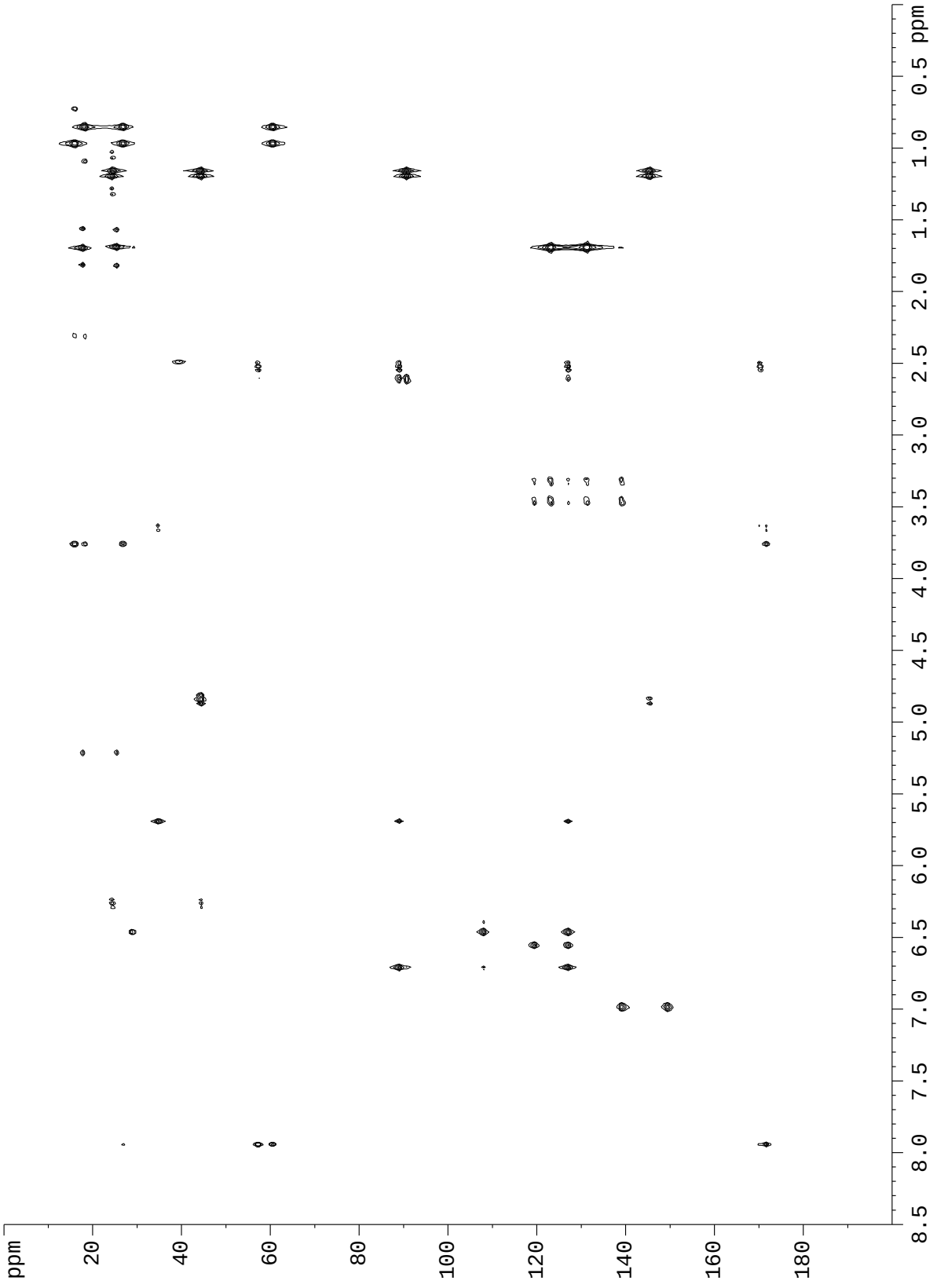
DQF-COSY


ppm
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5
ppm
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5

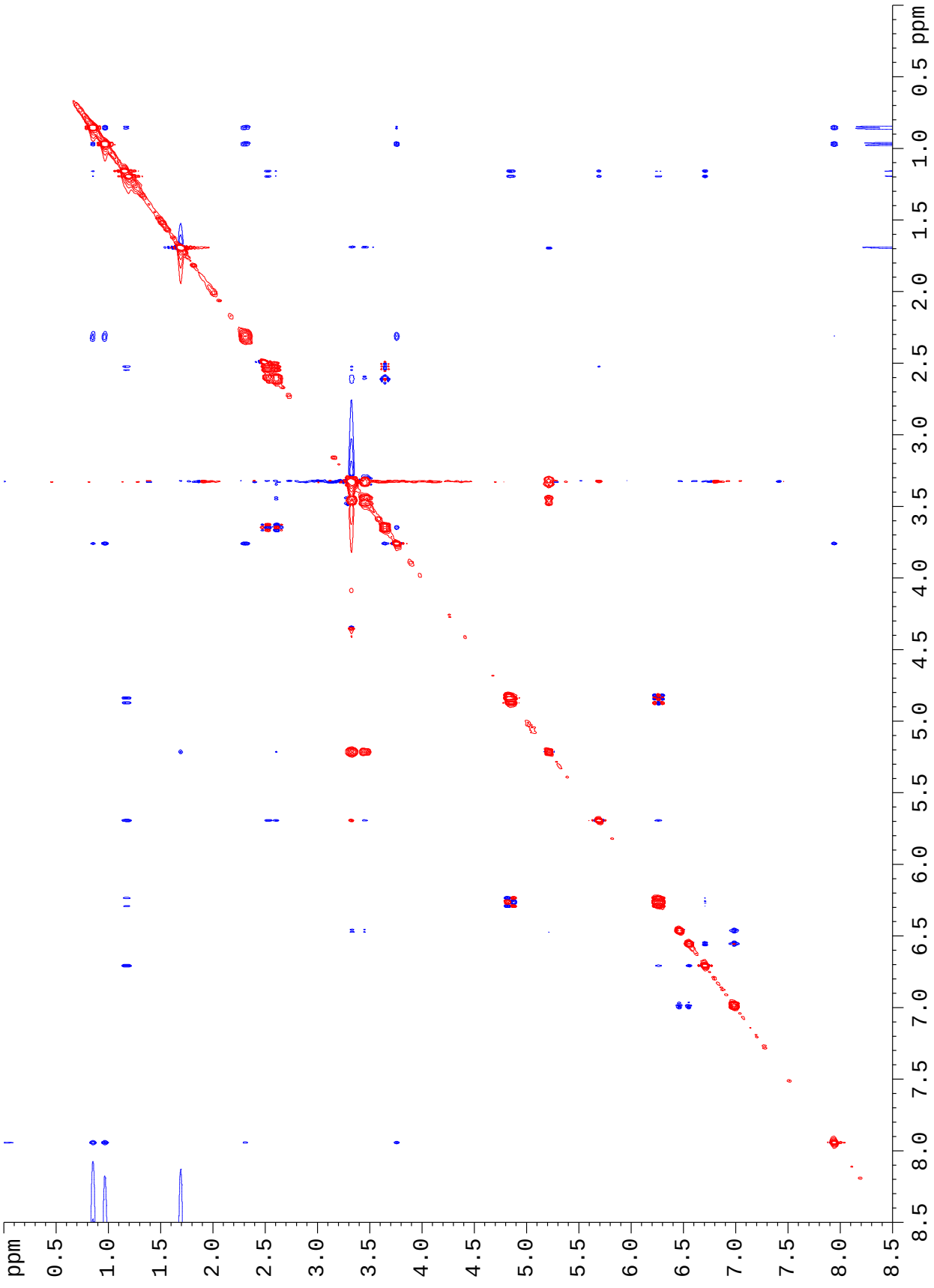
gHSQC


ppm
20
40
60
80
100
120
140
160
180
200
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
ppm

gHMBC


ppm
20
40
60
80
100
120
140
160
180
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
ppm

ROESY (150 ms)



*Marfey's reagents*

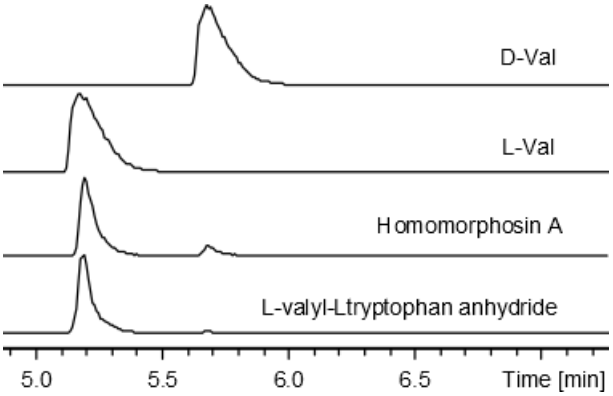



**Figure S1.** The retention times of the FDAA derivatives compared with retention times of the standard amino acid derivatives.

*UV-data*

The UV data shows that **1**, **2**, **3** and **6** probably share a similar chromophore. This is different than the chromophore of **4**, **5** and **6** which corresponds to the absorption of an indole ring.

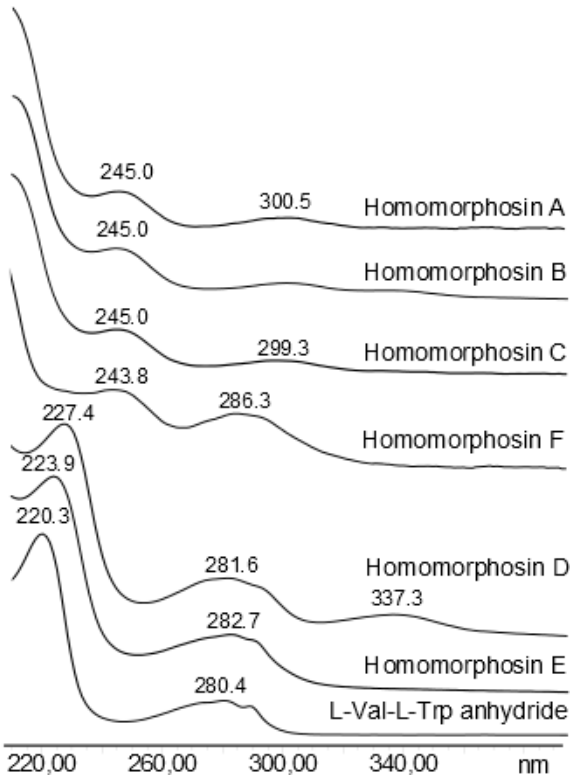



**Figure S2.** UV data of the homomorphosins.

*Stereochemistry of homomorphosin A*

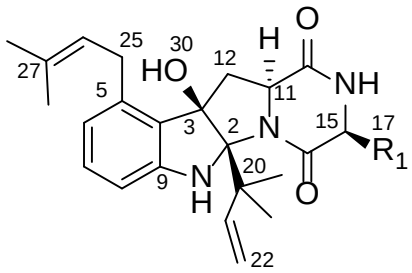



**Figure S3.** Numbering of homomorphosin A

**Table S1.** Distances of homomorphosin A from ISPA in Å. a: used as reference.

| Nucleus 1 | Nucleus 2 | Exp Dist | Lower bound | Violation | Upper bound | Violation |
|---|---|---|---|---|---|---|
| 1 | 8 | 2.66 | 2.39 | | 2.92 | |
| 1 | 21 | 2.88 | 4.18 | 1.30 | 5.11 | |
| 1 | 23/24 | 2.44 | 2.10 | | 2.56 | |
| 6 | 25a | 2.78 | 2.70 | | 3.30 | |
| 6 | 26 | 2.92 | 2.71 | | 3.31 | |
| 6 | 29 | 3.67 | 3.32 | | 4.05 | |
| 7 | 8 | 2.48[a] | 2.23 | | 2.73 | |
| 11 | 15 | 2.46 | 2.20 | | 2.69 | |
| 14 | 15 | 2.72 | 2.46 | | 3.00 | |
| 14 | 17 | 2.94 | 3.33 | 0.39 | 4.07 | |
| 15 | 18 | 3.05 | 3.17 | 0.12 | 3.88 | |
| 15 | 19 | 2.58 | 2.25 | | 2.75 | |
| 15 | 17 | 2.53 | 2.27 | | 2.77 | |
| 17 | 18 | 2.30 | 2.08 | | 2.54 | |
| 17 | 19 | 2.30 | 2.07 | | 2.53 | |
| 21 | 23/24 | 2.47 | 2.26 | | 2.77 | |
| 21 | 12a | 2.87 | 2.44 | | 2.98 | |
| 21 | 30 | 2.80 | 2.96 | 0.16 | 3.62 | |
| 22b | 23/24 | 2.28 | 2.07 | | 2.53 | |
| 22b | 18 | 4.30 | 3.69 | | 4.51 | |
| 25b | 29 | 2.59 | 2.46 | | 3.00 | |
| 25b | 12b | 2.26 | 1.88 | | 2.30 | |
| 26 | 28 | 2.39 | 2.15 | | 2.63 | |
| 30 | 23/24 | 2.41 | 2.38 | | 2.91 | |
| 30 | 12a | 2.84 | 1.92 | | 2.35 | 0.49 |
| 30 | 25b | 2.95 | 2.89 | | 3.53 | |

**Table S2.** Coupling *J*-constants in Hz for homomorphosin A.

| H1 | H2 | *J* exp. | *J* calc | Diff |
|---|---|---|---|---|
| 11 | 12b | 7.3 | 8.0 | 0.7 |
| 15 | 17 | 2.0 | 2.3 | 0.3 |
| 11 | 12a | 11.5 | 10.3 | 1.2 |

When averaging over the 364 structures from a conformational search, and only taking the structures into account which leads to increased fit iteratively over 10,000 steps, the errors are all but eliminated as seen below. This serves to show that the errors seen for a single structure may be partially due to flexibility and subsequent conformational averaging.

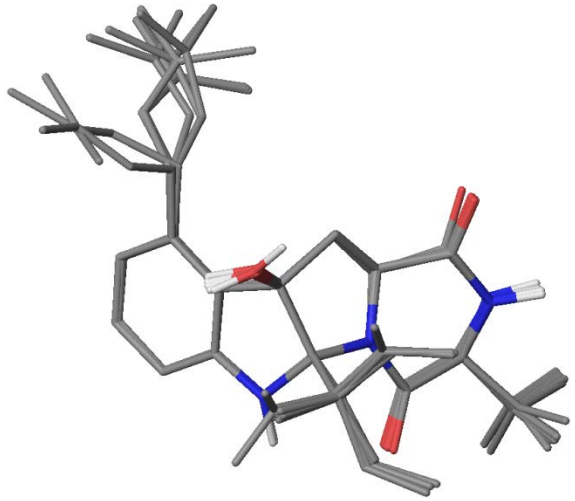


**Figure S4.** Superposition of the cyclic heavy atoms of ten structures which give the lowest averaged error.

**Table S3.** Average distances for homomorphosin A from ISPA in Å. a: used as reference.

| Nucleus 1 | Nucleus 2 | Experimental distance | Averaged distance | Diff (Å) | Diff (%) |
|---|---|---|---|---|---|
| **1** | **8** | **2.61** | **2.66** | **0.05** | **1.9** |
| 1 | 21 | 2.88 | 2.96 | 0.08 | 2.7 |
| 1 | 23/24 | 3.34 | 3.37 | 0.03 | 0.9 |
| 6 | 25a | 2.78 | 2.76 | 0.03 | 0.9 |
| 6 | 26 | 2.79 | 2.80 | 0.01 | 0.3 |
| 6 | 29 | 4.41 | 4.42 | 0.01 | 0.1 |
| **7** | **8** | **2.48**[a] | **2.48** | **0.00** | **0.0** |
| **11** | **15** | **2.53** | **2.44** | **0.09** | **3.7** |
| **14** | **15** | **2.72** | **2.73** | **0.00** | **0.1** |
| 14 | 17 | 2.94 | 2.93 | 0.01 | 0.5 |
| 15 | 18 | 3.42 | 3.50 | 0.08 | 2.3 |
| 15 | 17 | 2.25 | 2.52 | 0.27 | 12.1 |
| 15 | 19 | 2.76 | 3.01 | 0.25 | 9.1 |
| **17** | **18** | **2.81** | **2.62** | **0.20** | **7.0** |
| **17** | **19** | **2.75** | **2.61** | **0.14** | **5.0** |
| 21 | 12a | 2.87 | 2.96 | 0.09 | 3.1 |
| 21 | 23/24 | 3.10 | 3.28 | 0.18 | 5.8 |
| 21 | 30 | 2.77 | 2.89 | 0.12 | 4.2 |
| 22b | 18 | 4.71 | 4.67 | 0.04 | 0.8 |
| 22b | 23/24 | 3.05 | 3.16 | 0.11 | 3.6 |
| 25b | 12b | 2.18 | 2.27 | 0.09 | 4.1 |
| 25b | 29 | 3.11 | 3.07 | 0.04 | 1.3 |
| **26** | **28** | **2.96** | **2.82** | **0.14** | **4.6** |
| 30 | 12a | 2.84 | 2.75 | 0.09 | 3.1 |
| 30 | 23/24 | 3.09 | 3.35 | 0.26 | 8.5 |
| 30 | 25b | 2.81 | 2.92 | 0.11 | 4.0 |

**Table S4.** Average *J*-coupling constants in Hz for homomorphosin A calculated by the HLA, or Altona, equation.

| H1 | H2 | *J* HLA | *J* meas. | Diff |
|---|---|---|---|---|
| 11 | 12b | 6.5 | 7.3 | 0.8 |
| 11 | 12a | 10.2 | 11.5 | 1.3 |
| 15 | 17 | 2.0 | 2.0 | 0.0 |

*DP4 investigation of homomorphosin C*

A DP4 statistical investigation was conducted as suggested by Smith and Goodman as no NOE was observed between H11 and H15 for homomorphosin C, possibly due to very low amounts of compound isolated.[32] Homomorphosin C was truncated at C-20 and C-25 to simplify due to flexibility in the chains. Methyl groups were used as replacements. The shielding of C-2 and C-5 was not used in the DP4 calculation, nor was any protons attached to heteroatoms. This resulted in eight proton chemical shifts and twelve carbons. The four diastereomers were optimized to B3LYP/6-31G(d) level and shielding tensors were calculated using mPW1PW91/6-311+G(2d,p) and GAIO in Gaussian09, revision C.01 (Gaussian, Inc., Wallingford CT, 2013). The IEFPCM solvent model was used for DMSO. Tables with calculated and experimental chemical shifts and the result of the DP4 analysis are included below.

**Table S5.** Experimental vs. calculated $^{1}$H shifts in ppm. The stereochemistry of 1 and 14 is given as 1,14.

| H | Exp. | L,L | Δ | D,L | Δ |
|---|---|---|---|---|---|
| 6 | 6.57 | 6.50 | 0.07 | 6.59 | 0.02 |
| 7 | 6.99 | 7.03 | 0.04 | 7.10 | 0.11 |
| 8 | 6.47 | 6.35 | 0.12 | 6.41 | 0.06 |
| 11 | 3.72 | 3.85 | 0.13 | 3.81 | 0.09 |
| 15 | 3.89 | 3.95 | 0.06 | 3.80 | 0.09 |
| 17 | 1.16 | 1.08 | 0.08 | 1.24 | 0.08 |
| 12a | 2.63 | 2.98 | 0.35 | 2.82 | 0.19 |
| 12b | 2.53 | 2.38 | 0.15 | 2.03 | 0.50 |

**Table S5.** Experimental vs. calculated $^{13}$C shifts in ppm. The stereochemistry of 1 and 14 is given as 1,14.

| C | Exp. | L,L | Δ | D,L | Δ |
|---|---|---|---|---|---|
| 3 | 89.0 | 87.8 | 1.2 | 87.0 | 2.0 |
| 4 | 127.2 | 126.9 | 0.3 | 126.7 | 0.5 |
| 6 | 119.6 | 120.1 | 0.5 | 120.2 | 0.6 |
| 7 | 129.5 | 131.0 | 1.5 | 131.2 | 1.7 |
| 8 | 108.2 | 106.7 | 1.5 | 106.3 | 1.9 |
| 9 | 149.4 | 151.1 | 1.7 | 151.8 | 2.4 |
| 11 | 57.4 | 58.9 | 1.5 | 55.8 | 1.6 |
| 12 | 34.6 | 33.6 | 1.0 | 36.9 | 2.3 |
| 13 | 169.7 | 169.8 | 0.1 | 171.8 | 2.1 |
| 15 | 51.3 | 53.3 | 2.0 | 55.3 | 4.0 |
| 16 | 172.8 | 170.6 | 2.2 | 170.9 | 1.9 |
| 17 | 14.9 | 13.6 | 1.3 | 13.3 | 1.6 |

**Table S5.** The result of the DP4. The stereochemistry of 1 and 14 is given as 1,14.

| DP4 | L,L | D,L |
|---|---|---|
| | 97.7 | 2.3 |

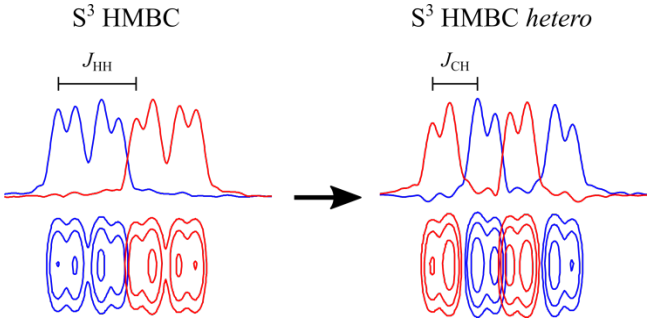
# $S^3$ HMBC *hetero*: Spin-State-Selective HMBC for accurate measurement of long-range *heteronuclear* coupling constants.

Casper Hoeck, Charlotte H. Gotfredsen* and Ole W. Sørensen*

Department of Chemistry, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

* Corresponding authors:
Charlotte H. Gotfredsen, E-mail:chg@kemi.dtu.dk, Tel.: +45 4525 2148
Ole W. Sørensen, E-mail: sorensen.ole.w@gmail.com

## Abstract




A novel method, Spin-State-Selective ($S^3$) HMBC *hetero*, for accurate measurement of heteronuclear coupling constants is introduced. The method extends the $S^3$ HMBC technique for measurement of homonuclear coupling constants by appending a pulse sequence element that interchanges the polarization in e.g. $^{13}C$-$^{1}H$ pairs. This amounts to converting the spin-state selectivity from $^{1}H$ spin states to $^{13}C$ spin states in the spectra of long-range coupled $^{1}H$ spins, allowing convenient measurement of heteronuclear coupling constants similarly to other $S^3$ or E.COSY-type methods. As usual in this type of techniques, the accuracy of coupling constant measurement is independent of the size of the coupling constant of interest. The merits of the new method are demonstrated by application to vinyl acetate, the alkaloid strychnine, and the carbohydrate methyl $\beta$-maltoside.

## Keywords

$S^3$ HMBC; HMBC; RDC; *J* coupling constants; heteronuclear; small molecules

## Introduction

Ever since the first discoveries of correlations between *J* coupling constants (*J*s) and molecular structure, methods for determining *J*s has been of keen interest .[1,2] Both the magnitude and the sign of *J*s are necessary for correlation with structure, thus favoring techniques where both the value and sign of the coupling

constant is determined, unless the sign is known in advance. More recently, residual dipolar coupling constants (RDCs) have become a valuable addition to *J* coupling constants in structural work, and as RDCs depend on a time averaged molecular orientation in a suitable alignment medium, their signs are crucial in obtaining structural information.[3,4] For small organic molecules the most abundant RDCs determined are $^1D_{CH}$ in weak alignment media where $^1D_{CH} < ^1J_{CH}$, and the sign is always known. This is not necessarily the case for long range coupling constants and thus the sign of $^nJ$ and $^nD$ is of utmost importance.

The probably most prominent example of correlation between *J* and structure is represented by $^3J_{HH}$ coupling constants and dihedral angles,[5–7] but further information is available from long range heteronuclear *J*s and RDCs, in particular in small organic molecules, which is the subject of this Communication.

It extends the recently published technique $S^3$ HMBC (Spin-State-Selective HMBC) designed for accurate measurement of homonuclear coupling constants based on editing of an HMBC-type spectrum into two subspectra corresponding to the α and β spin states of the $^1$H spin in a $^{13}$C-$^1$H methine group, which is long-range coupled to another proton. The extension results in the conversion of the $^1$H spin-state selectivity to a $^{13}$C spin-state selectivity by application of a zero-quantum (ZQ) π rotation within the $^{13}$C-$^1$H methine group. A similar approach was used in connection with the original $S^3$ work.[8]We dub the new experiment $S^3$ HMBC *hetero* and note that it is only applicable to $^{13}$C methine groups long-range coupled to protons of any kind.

**Method**

The approach to the new experiment is simplistic in that it includes the same five pulse sequences as $S^3$ HMBC, with the ZQ π rotation inserted right before acquisition sequences. The resulting data may thus be processed identically for both $S^3$ HMBC experiments. Therefore it suffices to briefly review the principles of $S^3$ HMBC.

The five pulse sequences which constitute the experiment include three from multiplicity-edited HMBC[13] and two from the HAT HMBC experiment that also includes multiplicity editing.[9,10] In the latter two experiments there are an additional $S^3$ $\pi_z(S)$ rotation that creates an antiphase relationship between the resonances associated with the α and β spin states of the passive $^1$H spin attached to $^{13}$C nuclei, which is the basis of the original $S^3$ HMBC. All five pulse sequences of $S^3$ HMBC *hetero* are outlined in Fig. 1.

The two subspectra for extraction of the coupling constants using $S^3$ HMBC *hetero* are generated by the following two linear combinations of time-domain data:

$$\left\{[a] - \tfrac{1}{2}[b + c]\right\} \pm \{d + e\} \quad (1)$$

where *a* to *e* refer to the individual pulse sequences in Fig. 1. The two resulting data sets are processed as ordinary HMBC data in magnitude mode and the subspectra contain methine as well as methyl group $^{13}$C cross peaks. Subspectra containing quaternary and methylene $^{13}$C cross peaks may be generated by other linear combinations, but they are not relevant in the present context.

The high editing accuracy of the $S^3$ HMBC experiments is caused by elimination of error terms in the [*b*+*c*] and {*d*+*e*} linear combination which, except for perfect matching of $\tau = (2^1J)^{-1}$, results in slightly reduced signal intensity. This loss can for the sake of cross talk reduction be recovered by scaling [*b*+*c*] and {*d*+*e*} by small factors *k* and $(2k/[k+1])^{½}$, respectively. These factors correspond to $-1/\cos(2\pi J\tau)$ and $1/\sin(\pi J\tau)$, respectively, for methine groups. It is an option to apply the correction to individual cross peaks, but typically $k = 1.05$ is employed throughout the entire spectral data set.

To illustrate the relationship between $S^3$ HMBC and $S^3$ HMBC *hetero*, spectra of vinyl acetate from the two experiments are shown in Fig. 2. In $S^3$ HMBC the displacement of the multiplets in the two subspectra correspond to $^3J$(H3,H4a) and $^3J$(H3,H4b), respectively, whilst in $S^3$ HMBC *hetero* the displacement of the multiplets in the two subspectra correspond to $^2J$(C3,H4a) and $^2J$(C3,H4b), respectively. The difference in multiplet patterns in the subspectra reflects the interchange of passive $^1$H3 and $^{13}$C3 spin states between $S^3$ HMBC and $S^3$ HMBC *hetero*. Literature values for the two $^2J$(C3-H4a/b) coupling constants are +7.6 Hz and -7.9 Hz, respectively, which fit well to the extracted values.[11]

A little cross talk is visible in the C3-H4b cross peak. It can be eliminated by the procedure outlined elsewhere,[12] shown in the lower panel of Fig. 2. The resulting change in the measured coupling constant is only a minor, below 0.1 Hz .

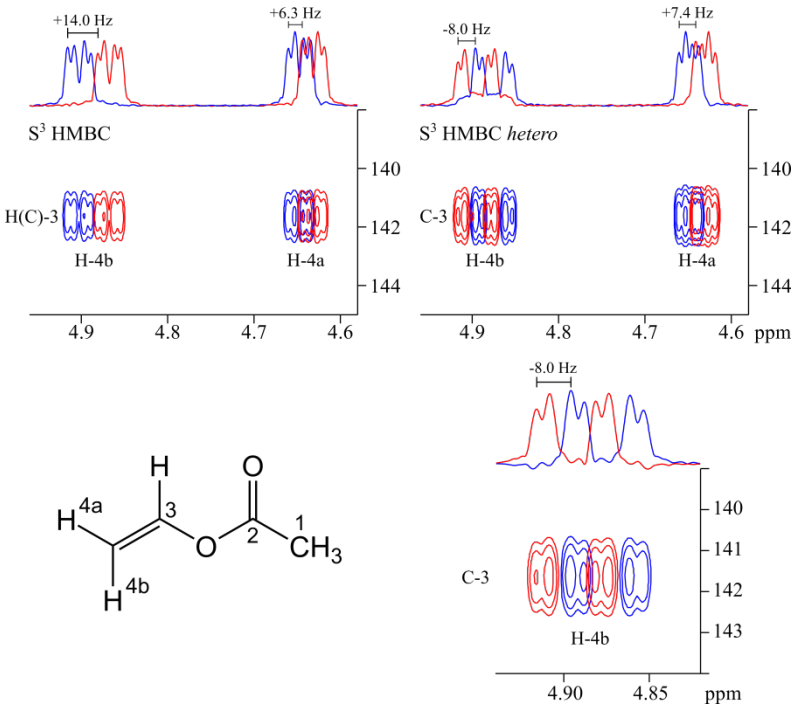



**Figure 1. Excerpts of cross peaks of vinyl acetate from $S^3$ HMBC and $S^3$ HMBC *hetero* spectra using identical setup (0.22 M in DMSO-$d_6$, 500 µl, 400 MHz, 16 scans, 4096 data points, 256 increments, zero filled to 32k × 1k, $J_{min}$ = 130 Hz, $J_{max}$ = 190 Hz, Δ = 62.5 ms). All subspectra were scaled to the same peak intensity. The spectra in the upper part are shown without further correction whilst the cross peak of C3-H4b in the lower part was corrected subtracting 16 % of the β subspectrum from the α subspectrum and vice versa to fully eliminate cross talk.[12]**

The additional ZQ π rotation in $S^3$ HMBC *hetero* is expected to cause only a minor reduction in sensitivity compared to $S^3$ HMBC, which was confirmed by our experiments. Nevertheless, it is clear that the interchange of $^1H$ and $^{13}C$ spin-state selectivity critically depends on absence of large off-resonance effect on $^{13}C$. The MBOB or broadband HMBC approach to cover a broad range of heteronuclear long-range coupling constants can be readily employed by repeating the experiment using different Δ delays in the pulse sequences and co-adding the resulting edited subspectra.[13,14]

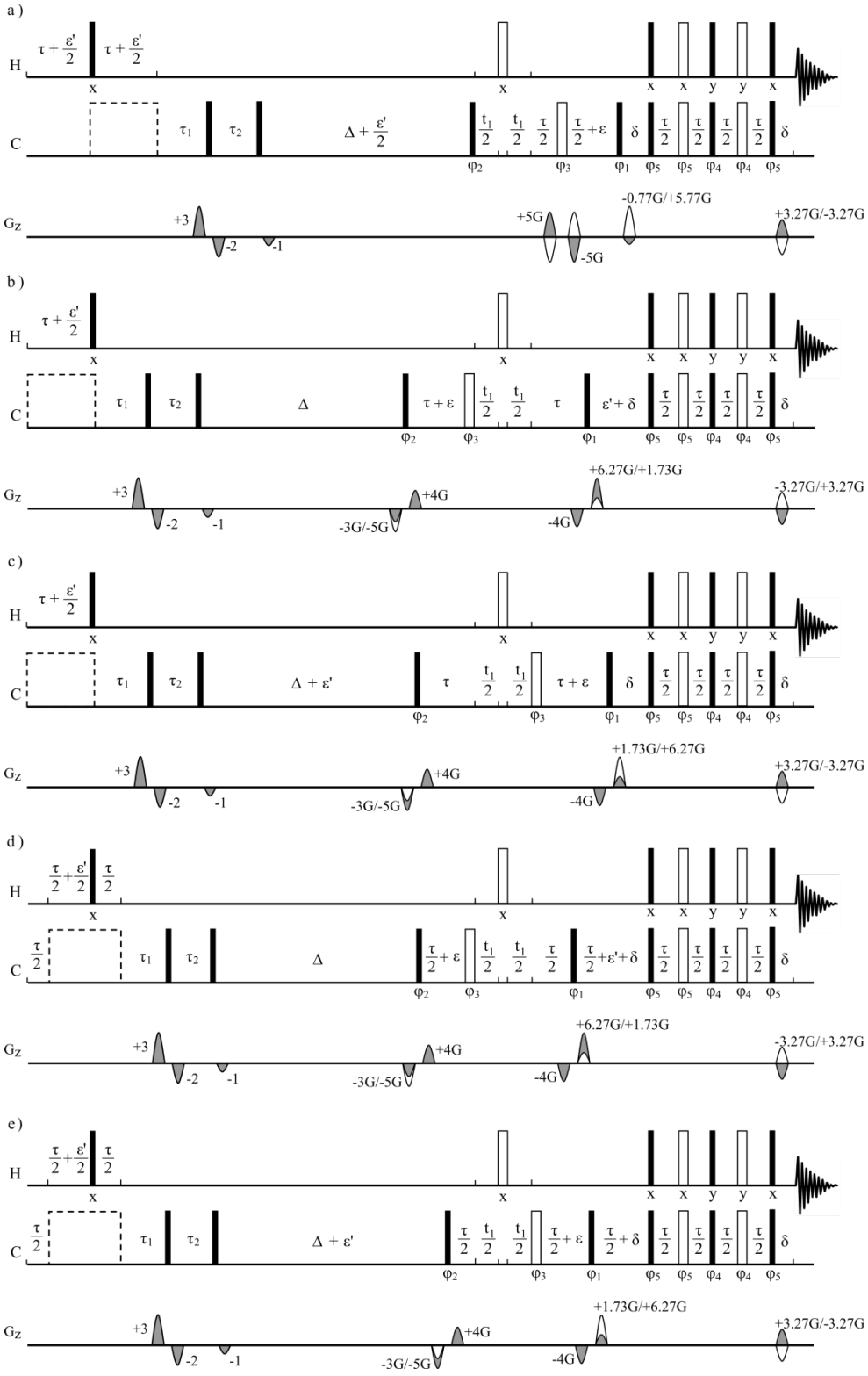



**Figure 2. S$^3$ HMBC *hetero* pulse sequence based on a-c) the sequences for edited HMBC, and d,e) the sequences for HAT HMBC, all shown with a 2$^{nd}$ order low-pass *J* filter. Filled and open bars refer to $\pi/2$ and $\pi$ pulses, the dashed open boxes represent $^{13}$C decoupling. $\tau = (2\ ^1J_{CH})^{-1}$ or $(^1J_{max} + {}^1J_{min})^{-1}$, $\delta$ is a gradient delay, $\varepsilon = 2\ t_{1/2,min} + t(\pi_H)$, $\varepsilon' = \varepsilon + t(\pi_C)$, $\tau_1 = \frac{1}{2}[^1J_{min} + 0.146\ (^1J_{max} - {}^1J_{min})]^{-1}$, $\tau_3 = \frac{1}{2}[^1J_{max} - 0.146\ (^1J_{max} - {}^1J_{min})]^{-1}$. $\Delta$ is the delay for evolution under heteronuclear long-range coupling constants. The gradient pair encasing the new pulses was set to $\pm 4/\sqrt{1.5}$G ~ ±3.27G . $\varphi_1$ = {x, x 4(-x), x, x}, $\varphi_2$ = {4(x), 8(-x), 4(x)}, $\varphi_3$ = {8(x), 8(y), 8(-x), 8(-y)}, $\varphi_4$ = {y, -y, -y, y}, $\varphi_5$ = {x, -x, -x, x}.**

## Application

To demonstrate the $S^3$ HMBC *hetero* pulse sequence, $^{n}J_{CH}$ coupling constants of the alkaloid strychnine were determined. Strychnine is the benchmark molecule for heteronuclear small-molecule NMR experiments, and a large number of $^{n}J_{CH}$ coupling constants have been previously published for comparison, see Table 1.[15–19] As for $S^3$ HMBC, the sizes and signs of the coupling constants are easily extracted by measuring the multiplet displacements in the two subspectra, leading to the determination of 23 $^{n}J_{CH}$ coupling constants. Of these, 20 were included in the chosen comparison publications and are in excellent agreement. Most of the comparative experiments do not yield the sign of the coupling constants, but in all those that do there is full agreement with the signs determined with $S^3$ HMBC *hetero*. All signs determined also agree with theoretical coupling constants from DFT calculations. The calculated theoretical coupling constants may be scaled, as previously shown for other coupling constant calculations, to obtain a better correlation to the data.[19,20] Excerpts of representative cross peaks are found in Figure 3.

Coupling constants were calculated in Gaussian09 using MPW1PW91/6-311G+(d,p) from the lowest energy structures optimized to the same level of theory which was found to give a good correlation in between experimental and theoretical data. The theoretical coupling constants are in line with previously reported calculated coupling constants of strychnine.[19,21]

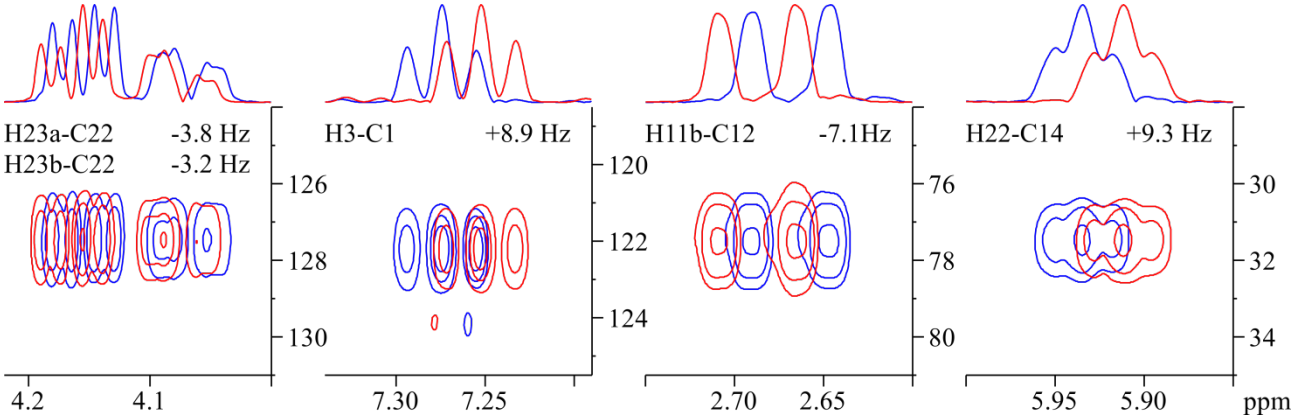
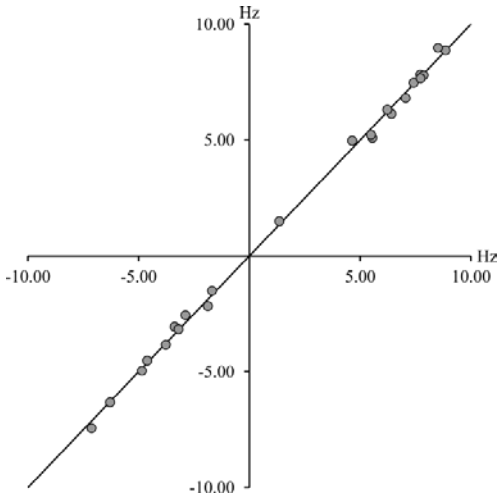



**Figure 3. Examples $S^3$ HMBC *hetero* cross peaks for strychnine (0.18 M in $CDCl_3$, 500 µl, 400 MHz, 16 scans, 4096 data points, 256 increments, zero filled to 32k × 1k, $J_{min}$ = 124 Hz, $J_{max}$ = 167 Hz, δ = 62.5 ms). The coupling constants were measured by displacing 1D sections in analogy to the E.COSY method. The excerpts were scaled individually to the same intensity.**

**Table 1. Experimental and theoretical $^{n}J_{CH}$ coupling constants of strychnine measured at 400 MHz, compared to experimental data from different experiments, see references [15–18]. Both optimization and *J* coupling constant calculations were performed at a MPW1PW91/6-311+G(d,P) level of theory. Calculated coupling constants are given without scaling and linearly scaled to the experimental data as $J_{scaled} = (J_{calc}-b)/a$, with $a = 0.92$, $b = -0.14$. The figure on the right depicts $S^3$ HMBC *hetero* data versus the scaled theoretical data, RMSD = 0.23 Hz.**

| H | C | Experimental (Hz) | | | | | Theoretical (Hz) | |
|---|---|---|---|---|---|---|---|---|
| # | # | S3 HMBC *hetero* | [15] | [16] | [17] | [18] | Calc. | Scaled |
| 1 | 3 | 7.4 | 7.5,7.2 | | | | 6.7 | 7.4 |
| 2 | 4 | 7.9 | | | | | 7.0 | 7.8 |
| 3 | 1 | 8.9 | 8.9,9.3 | | | | 8.1 | 8.9 |
| 4 | 2 | 7.7 | | 7.5,7.4 | 7.9 | | 7.1 | 7.8 |
| 8 | 12 | 6.4 | 5.8,5.6 | 5.5,5.6 | 5.3 | | 5.6 | 6.2 |
| 8 | 13 | -1.7 | | | | | -1.7 | -1.6 |
| 11b | 12 | -7.1 | 6.9,-6.9 | 7,6.9 | 6.8 | | -6.9 | -7.3 |
| 13 | 8 | -6.3 | | 6.3,6.4,6.3 | 6.2 | | -6.0 | -6.3 |
| 13 | 14 | -4.9 | 5.5,-4.5 | 4.6,4.7 | 4.5 | 4.7 | -4.7 | -5.0 |
| 15a | 13 | 7.7 | | 8,8.1 | 8 | 7.9 | 7.0 | 7.7 |
| 15a | 14 | -2.9 | 1.2,-2.9 | 1.8,2.8,2.8 | 2 | 3.2 | -2.5 | -2.6 |
| 15a | 16 | -4.6 | | 4.5,4.3 | 4 | | -4.3 | -4.5 |
| 15b | 14 | -3.2 | 2.3,-2.6 | 3.2,3.6 | 2.8 | 4.8 | -3.1 | -3.2 |
| 15b | 16 | -1.9 | | 2.7,3 | 1.4 | | -2.1 | -2.2 |
| 16 | 14 | 6.2 | 6.7,6.4 | | | 6.4 | 5.8 | 6.4 |
| 20a | 22 | 1.4 | | | | | 1.3 | 1.6 |
| 20a | 22 | 5.6 | 4.7,5 | 6.1,5.8 | 5.1 | | 4.6 | 5.1 |
| 20b | 14 | 5.5 | 5.7,5.4 | 5.4,5.3 | 5.1 | 5.5 | 4.7 | 5.2 |
| 20b | 16 | 7.8 | | 6.9,6.9 | 6.7 | | 6.3 | 6.9 |
| 20b | 22 | 4.6 | 5.1,5.5 | 4.5,4.6 | 4.2 | | 4.6 | 5.1 |
| 22 | 14 | 8.5 | 8.5,8.6 | 7.9,8.5,8.6 | 8.8 | 8.9 | 8.0 | 8.8 |
| 23a | 22 | -3.8 | 4.2,-3.8 | 3.8,4 | | | -3.8 | -3.9 |
| 23b | 22 | -3.4 | 3.6,-3.4 | 4,3.9 | | | -3.0 | -3.1 |




An immediate application of the new pulse sequence is in conformational studies of polysaccharides, where the heteronuclear coupling constants across the glycosidic linkage have proven valuable.[22–25] To test the applicability of the $S^3$ HMBC hetero in the extraction of such coupling constants, spectra of the disaccharide methyl $\beta$-maltoside were acquired, and a coupling constant across the glycosidic linkage is shown in Figure 4. The $^{3}J_{COCH}$-coupling constant may be compared to a previously reported value of 4.0 Hz.[25] The $^{3}J_{CH}$-coupling constant of the C1'-H4 pair was not determined, due to spectral overlap in between H4 and H2' and strong coupling interference from the strongly coupled H4-H5 proton pair.

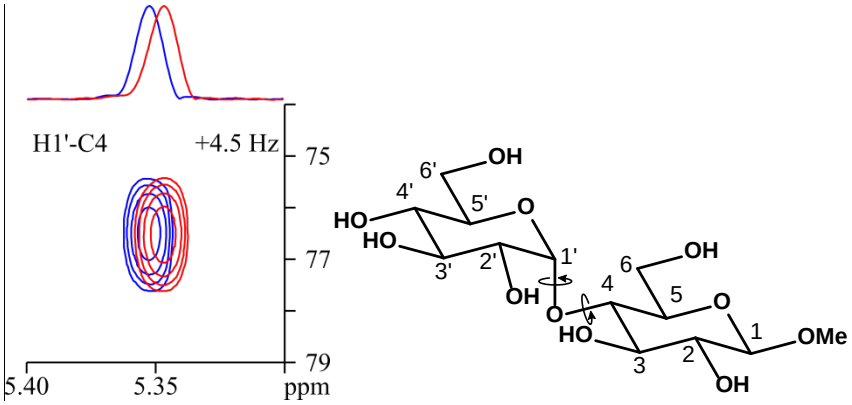



**Figure 4. A $S^3$ HMBC *hetero* cross peak across the glycosidic linkage of methyl *β*-maltoside, shown to the right. (56 mM in $D_2O$, 500 µl, 800 MHz, 16 scans, 4096 data points, 256 increments, zero filled to 16k × 1k, $J_{min}$ = 140 Hz, $J_{max}$ = 171 Hz, δ = 125 ms). The coupling constants were measured by displacing 1D sections in analogy to the E.COSY method.**

## Conclusions

The $S^3$ HMBC experiment for measurement of homonuclear coupling constants has been extended to cover heteronuclear coupling constants involving methine groups and dubbed $S^3$ HMBC *hetero*. The extension consists of insertion of a ZQ π rotation in the pulse sequence prior to acquisition in $S^3$ HMBC. The method was successfully tested on the alkaloid strychnine with accurate measurement of 23 long-range $J_{CH}$. The experiment is expected not least to find widespread application to carbohydrates where methine groups are abundant and the valuable conformational information contained in the inter-glycosidic $^3J_{CH}$ coupling constants, which was exemplified by the disaccharide methyl *β*-maltoside.[22,25,26]

**References**


(1) Parella, T.; Espinosa, J. F. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 17–55.

(2) Marquez, B. L.; Gerwick, W. H.; Williamson, R. T. *Magn. Reson. Chem.* **2001**, *39*, 499–530.

(3) Thiele, C. M. *Eur. J. Org. Chem.* **2008**, 5673–5685.

(4) Kummerlöwe, G.; Luy, B. *Trends Anal. Chem.* **2009**, *28*, 483–493.

(5) Haasnoot, C. A. G.; DeLeeuw, F. A. A. M.; Altona, C. *Tetrahedron* **1979**, *36*, 2783–2792.

(6) Karplus, S.; Karplus, M. *Proc. Natl. Acad. Sci. USA* **1972**, *69*, 3204–3206.

(7) Bifulco, G.; Dambruoso, P.; Gomez-Paloma, L.; Riccio, R. *Chem. Rev.* **2007**, *107*, 3744–3779.

(8) Meissner, A.; Duus, J. Ø.; Sørensen, O. W. *J. Biomol. NMR* **1997**, *10*, 89–94.

(9) Benie, A. J.; Sørensen, O. W. *Magn. Reson. Chem.* **2006**, *44*, 739–743.

(10) Benie, A. J.; Sørensen, O. W. *J. Magn. Reson.* **2007**, *184*, 315–321.

(11) Crecely, K. M.; Crecely, R. W.; Goldstein, J. H. *J. Mol. Spectrosc.* **1971**, *37*, 252–259.

(12) Sørensen, M. D.; Meissner, A.; Sørensen, O. W. *J. Magn. Reson.* **1999**, *137*, 237–242.

(13) Schulte-Herbrüggen, T.; Meissner, A.; Papanikos, A.; Meldal, M.; Sørensen, O. W. *J. Magn. Reson.* **2002**, *156*, 282–294.

(14) Meissner, A.; Sørensen, O. W. *Magn. Reson. Chem.* **2000**, *38*, 981–984.

(15) Gil, S.; Espinosa, J. F.; Parella, T. *J. Magn. Reson.* **2010**, *207*, 312–321.

(16) Edden, R. A. E.; Keeler, J. *J. Magn. Reson.* **2004**, *166*, 53–68.

(17) Butts, C. P.; Heise, B.; Tatolo, G. *Org. Lett.* **2012**, *14*, 3256–3259.

(18) Blechta, V.; del Río-Portilla, F.; Freeman, R. *Magn. Reson. Chem.* **1994**, *32*, 134–137.

(19) Bagno, A.; Rastrelli, F.; Saielli, G. *Chem. Eur. J.* **2006**, *12*, 5514–5525.

(20) Bally, T.; Rablen, P. R. *J. Org. Chem.* **2011**, *76*, 4818–4830.

(21) Bifulco, G.; Riccio, R.; Martin, G. E.; Buevich, A. V; Williamson, R. T. *Org. Lett.* **2013**, *15*, 654–657.

(22) Pendrill, R.; Säwén, E.; Widmalm, G. *J. Phys. Chem. B* **2013**, *117*, 14709–14722.

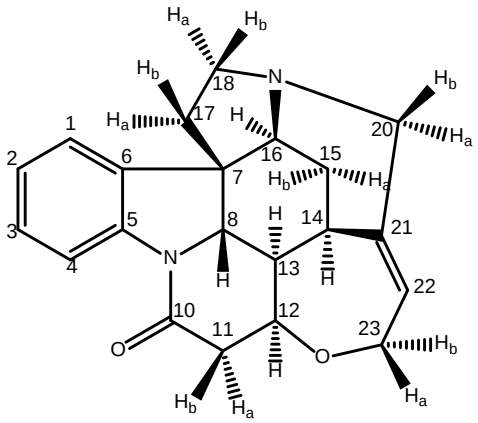
(23) Martin-Pastor, M.; Bush, C. A. *J. Biomol. NMR* **2001**, *19*, 125–139.

(24) Rundlöf, T.; Kjellberg, A.; Damberg, C.; Nishida, T.; Widmalm, G. *Magn. Reson. Chem.* **1998**, *36*, 839–847.

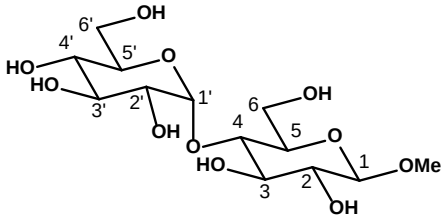
(25) Hatcher, E.; Säwén, E.; Widmalm, G.; Mackerell, Jr., A. D. *J. Phys. Chem. B* **2011**, *115*, 597–608.

(26) Otter, A.; Bundle, D. R. *J. Magn. Reson. Ser. B* **1995**, *109*, 194–201.

**Supporting Info**




| Atom # | $^{1}$H chemical shift [ppm] multiplicity, *J* [Hz] | $^{13}$C chemical shift [ppm] |
|---|---|---|
| 1 | 7.12, 1H, d, 7.6 | 122.3 |
| 2 | 7.05, 1H, t, 7.6, | 124.3 |
| 3 | 7.21, 1H, t, 7.8 | 128.6 |
| 4 | 8.05, 1H, d, 8.0 | 116.2 |
| 5 | - | 142.3 |
| 6 | - | 132.9 |
| 7 | - | 51.9 |
| 8 | 3.82, 1H, d, 10.4 | 60.1 |
| 9-N | - | - |
| 10 | - | 169.4 |
| 11a | 3.09, 1H, m | 42.4 |
| 11b | 2.63, 1H, dd, 17.5, 3.4 | 42.4 |
| 12 | 4.25, 1H, dt, 8.2, 3.4 | 77.5 |
| 13 | 1.23, 1H, dt, 10.5, 3.2 | 48.1 |
| 14 | 3.10, 1H, m | 31.5 |
| 15a | 2.31, 1H, dt, 14.4, 4.3 | 26.7 |
| 15b | 1.41, 1H, br. d, 14.5 | 26.7 |
| 16 | 3.90, 1H, br. S | 60.1 |
| 17a/b | 1.84, 2H, m | 42.8 |
| 18a | 3.15, 1H, m | 50.2 |
| 18b | 2.83, 1H, m | 50.2 |
| 19-N | - | - |
| 20a | 3.66, 1H, d, 14.9 | 52.6 |
| 20b | 2.69, 1H, d, 14.6 | 52.6 |
| 21 | - | 140.6 |
| 22 | 5.86, 1H, m | 127.3 |
| 23a | 4.10, 1H, dd, 13.7, 7.1 | 64.6 |
| 23b | 4.02, 1H, dd, 13.7, 5.9 | 64.6 |




| Atom # | $^{1}$H chemical shift [ppm] multiplicity, *J* [Hz] | $^{13}$C chemical shift [ppm] |
|---|---|---|
| 1 | 4.33, 1H, d, 8.0 | 102.9 |
| 2 | 3.23, 1H, dd, 9.4, 8.1 | 72.9 |
| 3 | 3.71, 1H, m | 76.1 |
| 4 | 3.55, 1H, m | 76.7 |
| 5 | 3.54, 1H, m | 74.4 |
| 6a | 3.72, 1H, m | 60.6 |
| 6b | 3.88, 1H, dd, 12.2, 1.8 | 60.6 |
| 1’ | 5.35, 1H, d, 4.1 | 99.4 |
| 2’ | 3.51, 1H, m | 71.8 |
| 3’ | 3.62, 1H, m | 72.6 |
| 4’ | 3.33, 1H, t, 9.5 | 69.2 |
| 5’ | 3.64, 1H, m | 72.8 |
| 6’a | 3.69, 1H, m | 60.3 |
| 6’b | 3.79, 1H, dd, 12.2, 2.1 | 60.3 |
| OMe | 3.50, 3H, s | 57.0 |

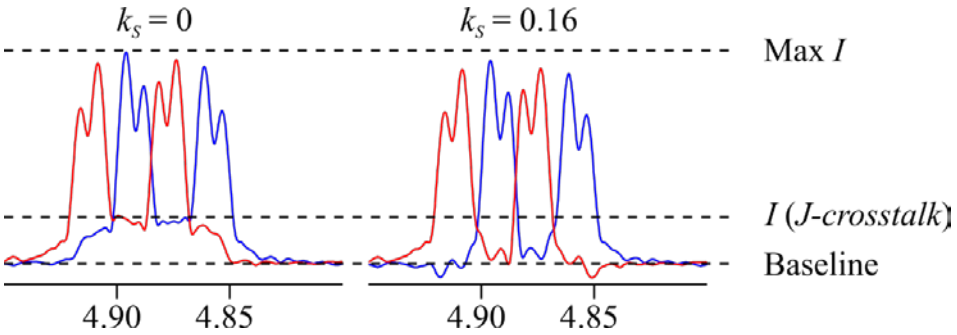



**Fig. S1. Comparison of the *J* cross talk in standard and edited spectra of the C3-H4a crosspeak of vinyl acetate. Edited as $\alpha_S = \alpha - k_S \cdot \beta$ and $\beta_S = \beta - k_S \cdot \alpha$. The difference in the extracted $J_{CH}$-coupling constant is below 0.1 Hz.**

Akin to the original S3 HMBC work, a script written in Matlab was used to extract the coupling constants from the subspectra. A comparison between this automated extraction, and the coupling constants determined manually is seen in Fig. S2

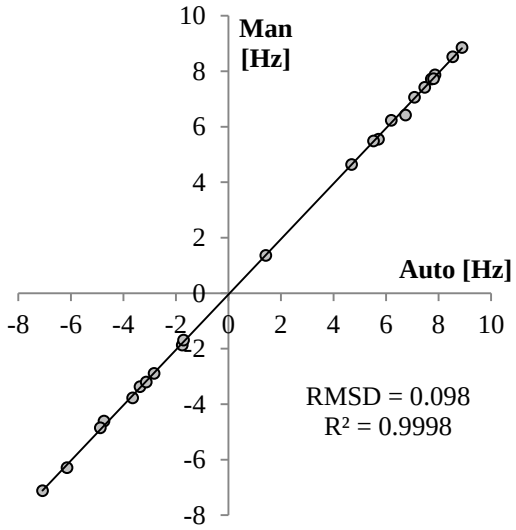



**Fig. S2. Manual versus automatic extraction of J coupling constants**

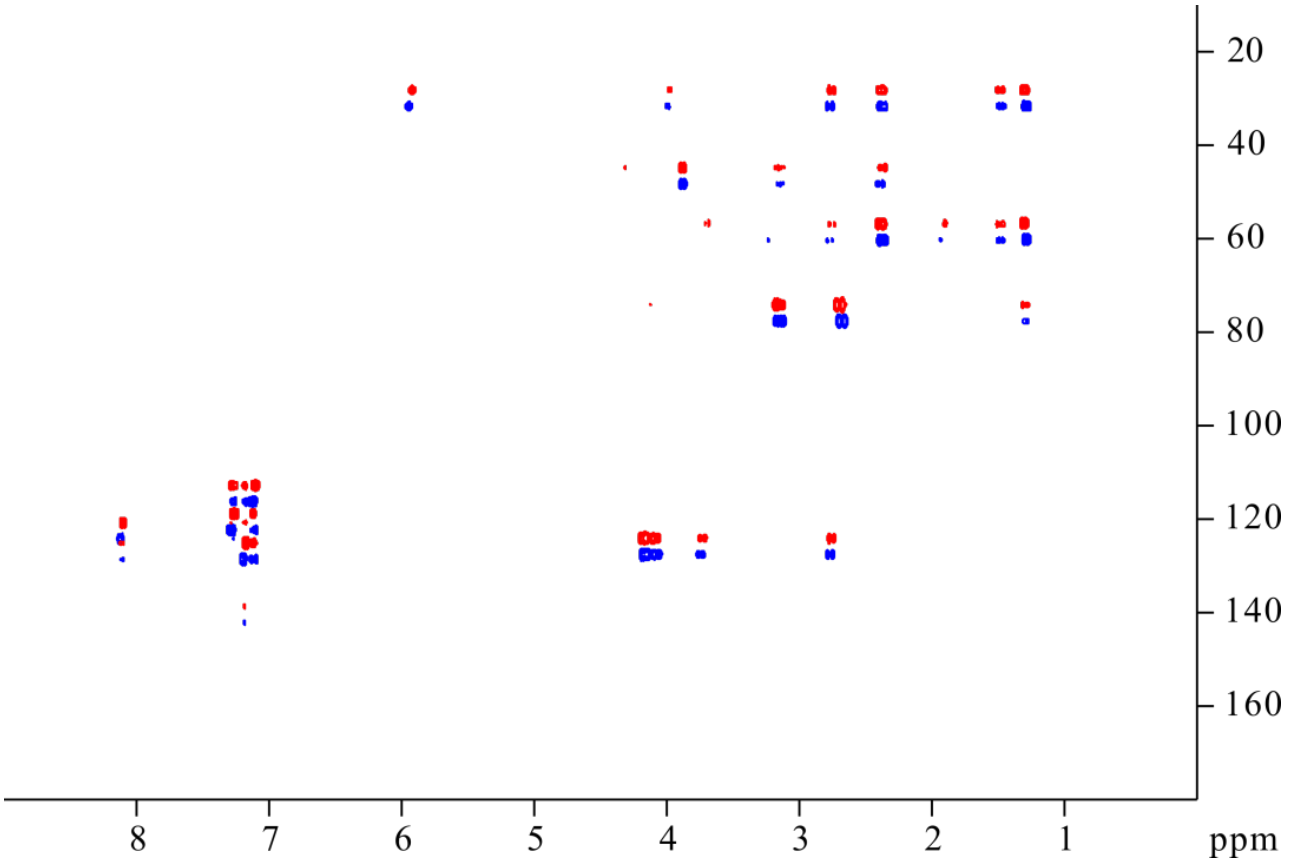



**Fig. S3. $S^3$ HMBC *hetero* spectrum of strychnine. (-) spectrum (red) offset vertically to better visualize displacements.**